HUMAN DEVELOPMENT

c

HUMAN DEVELOPMENT

Readings in research

Ira J. Gordon

UNIVERSITY OF FLORIDA

S.C.E.R.T., West Bengal
Library
Calcutta
25/3, B C. Road

D. B. TARAPOREVALA SONS & CO. PRIVATE LTD.
210 Dr. DADABHAI NAOROJI ROAD, BOMBAY 1

Copyright © 1965 by Scott, Foresman and Company,
Glenview, Illinois 60025. All rights reserved.

.E.R.T., West Bengal
te 23-2-72
c. No 2320

DEDICATION
To my wife, Esther,
and our children,
Gary and Bonnie

136.5
GOR

Copyright 1970 in India, Pakistan, Burma and Ceylon
by D. B. Taraporevala Sons & Co. Private Ltd.
by arrangement with Scott, Foresman and Company

First Indian Reprint, 1970

This book has been published with the assistance of the Joint Indian-American Textbook Programme. It includes every word and illustration contained in the latest original higher priced edition of this title and is a complete and authentic copy thereof.

Sales Territory: (1) All countries of Asia (except Formosa, Japan and the Philippine Islands)
(2) All countries of Africa (except South Africa and British Commonwealth countries in Africa)

This book can be exported from India only by the publishers, D. B. Taraporevala Sons & Co. Private Ltd. and this constitutes a condition of its initial sale and its subsequent sales. Infringement of this condition of sale will lead to civil and/or criminal legal action and prosecution.

Price in India only : Rs. 7.45

Printed and published by R. J. Taraporevala for D. B. Taraporevala Sons & Co. Private Ltd., 210, Dr. Dadabhai Naoroji Road, Bombay 1, at their Electrographic Industries Division, Apte Industrial Estate, Worli, Bombay 18.

ACKNOWLEDGMENTS

Gainesville, Florida
September 1964

Obviously, a book of readings owes most of its value to the original work of the many authors represented. I acknowledge my indebtedness to them, and to their publishers, for permission to reprint their material. Because of style and space considerations, I have deleted some of the general references, footnotes, tables, illustrations, and statistical material. Major deletions in these articles are indicated by ellipses. Where a table, chart, or photograph has been deleted, I have renumbered the remainder consecutively. None of these changes alter or distort the findings or theses of the authors in any way, and all major changes have received their approval.

The work of the editorial department of Scott, Foresman and Company deserves a special acknowledgment; its various members have rendered exceptional service to me in many ways. My thanks must also go to Mrs. Marguerite Robinson, my secretary, for her work in the preparation of this book.

I am also indebted to my students for their careful evaluation and reactions to many readings, which led to the selection of these particular articles.

No undertaking such as this would have been carried to completion without the support, patience, and understanding of my wife, Esther, and of the interest of Bonnie and Gary, to whom I affectionately dedicate this book.

Ira J. Gordon

Contents

INTRODUCTION: HOW TO READ THIS BOOK

PART I
GENERAL CONCEPTS

1. *J. McV. Hunt.* Experience and the Development of Motivation: Some Reinterpretations 3
2. *Robert W. White.* Motivation Reconsidered: The Concept of Competence 10
3. *J. P. Guilford.* Three Faces of Intellect 23
4. *Nancy Bayley.* Individual Patterns of Development 33
5. *Howard A. Moss & Jerome Kagan.* Stability of Achievement and Recognition Seeking Behaviors from Early Childhood Through Adulthood 47
6. *Jean Piaget.* The Genetic Approach to the Psychology of Thought 57
7. *David Elkind.* The Development of Quantitative Thinking: A Systematic Replication of Piaget's Studies 61
8. *Jerome S. Bruner.* The Act of Discovery 67
9. *David B. Lynn.* Sex-Role and Parental Identification 74

PART II
INFANCY

The Beginnings of Behavior

10. *William Kessen.* Research in the Psychological Development of Infants: An Overview 83
11. *William Kessen, E. Jane Williams, & Joanna P. Williams.* Selection and Test of Response Measures in the Study of the Human Newborn 90

12. *Hilda Knobloch & Benjamin Pasamanick.* The Developmental Behavioral Approach to the Neurologic Examination in Infancy 100

13. *Bettye M. Caldwell.* Assessment of Infant Personality 111

Maternal Behavior and Attitudes

14. *Harriet L. Rheingold.* The Measurement of Maternal Care 117

15. *Earl S. Schaefer & Richard Q. Bell.* Development of a Parental Attitude Research Instrument 123

The Mother-Child Relationship

16. *Niles Newton.* The Influence of the Let-Down Reflex in Breast Feeding on the Mother-Child Relationship 136

17. *Leone Kell & Joan Aldous.* Trends in Child Care over Three Generations 138

18. *Harry F. Harlow.* The Nature of Love 140

19. *Harry F. Harlow.* The Heterosexual Affectional System in Monkeys 150

PART III

EARLY CHILDHOOD

The Parent-Child Relationship

20. *Ethelyn H. Klatskin, Edith B. Jackson, & Louise C. Wilkin.* The Influence of Degree of Flexibility in Maternal Child Care Practices on Early Child Behavior 161

21. *Damaris Pease & Glenn R. Hawkes.* Observation and Parent-Child Interaction 170

22. *Harry H. L. Kitano.* Adjustment of Problem and Nonproblem Children to Specific Situations: A Study in Role Theory 174

Cognitive Development

23. *M. D. Vernon.* The Development of Perception in Children 177

24. *Ralph H. Ojemann & Karen Pritchett.* Piaget and the Role of Guided Experiences in Human Development 184

25. *Joachim F. Wohlwill & Roland C. Lowe.* Experimental Analysis of the Development of the Conservation of Number 193

Personality Development

26. *Marjorie Stith & Ruth Connor.* Dependency and Helpfulness in Young Children 202

27. *Joseph H. Meyerowitz.* Self-Derogations in Young Retardates and Special Class Placement 205

PART IV
MIDDLE CHILDHOOD

Parent-Child Relations

28. *Lois Wladis Hoffman.* Effects of Maternal Employment on the Child 211

29. *Vaughn J. Crandall & Anne Preston.* Verbally Expressed Needs and Overt Maternal Behaviors 217

30. *Walter Mischel.* Father-Absence and Delay of Gratification: Cross-Cultural Comparisons 223

The Child in School

31. *Jacob S. Kounin & Paul V. Gump.* The Comparative Influence of Punitive and Nonpunitive Teachers upon Children's Concepts of School Misconduct 231

32. *Ira J. Gordon & Patricia C. Wood.* The Relationship Between Pupil Self-Evaluation, Teacher Evaluation of the Pupil, and Scholastic Achievement 236

33. *Ralph H. Ojemann & Bill C. Snider.* The Effect of a Teaching Program in Behavioral Science on Changes in Causal Behavior Scores 240

34. *E. Paul Torrance & R. E. Myers.* Teaching Gifted Elementary Pupils Research Concepts and Skills 245

35. *Charlotte Antoinette Crabtree.* Effects of Structuring on Productiveness of Children's Thinking 249

36. *Mary W. Lamy.* Relationship of Self-Perceptions of Early Primary Children to Achievement in Reading 251

Peer Relations

37. *Marian Radke Yarrow & John D. Campbell.* Person Perception in Children 252

38. *Carlfred B. Broderick & Stanley E. Fowler.* New Patterns of Relationships Between the Sexes Among Preadolescents 261

Cognitive Development

39. *W. H. King.* The Development of Scientific Concepts in Children 265
40. *Leonore Boehm.* The Development of Conscience: A Comparison of American Children of Different Mental and Socioeconomic Levels 276
41. *Sonia F. Osler & Sandra Raynes Weiss.* Studies in Concept Attainment: III. Effect of Instructions at Two Levels of Intelligence 283
42. *Irving E. Sigel.* How Intelligence Tests Limit Understanding of Intelligence 287

Personality Development

43. *H. Harrison Clarke & David H. Clarke.* Relationship Between Level of Aspiration and Selected Physical Factors of Boys Aged Nine Years 297
44. *Rolf E. Muuss.* The Relationship between "Causal" Orientation, Anxiety, and Insecurity in Elementary School Children 300

The Impact of the Culture

45. *Harold H. Anderson & Gladys L. Anderson.* A Cross-National Study of Children: A Study in Creativity and Mental Health 307
46. *Wilbur Schramm, Jack Lyle, & Edwin B. Parker.* Children's Learning from Television 315

PART V
ADOLESCENCE

Physical Development

47. *Paul Henry Mussen & Mary Cover Jones.* Self-Conceptions, Motivations, and Interpersonal Attitudes of Late- and Early-Maturing Boys 325
48. *Mary Cover Jones & Paul Henry Mussen.* Self-Conceptions, Motivations, and Interpersonal Attitudes of Early- and Late-Maturing Girls 333

Family Influences

49. *Merville C. Shaw & Bert E. Dutton.* The Use of the Parent Attitude Research Inventory with the Parents of Bright Academic Underachievers 340
50. *Elias Tuma & Norman Livson.* Family Socioeconomic Status and Adolescent Attitudes to Authority 344

The Peer Society

51. *James S. Coleman.* The Adolescent Culture 351
52. *Hayne W. Reese.* Relationships Between Self-Acceptance and Sociometric Choices 363

Personality and Self

53. *Dorothy M. Kipnis.* Changes in Self Concepts in Relation to Perceptions of Others 366
54. *Earle Silber, David A. Hamburg, George V. Coelho, Elizabeth B. Murphey, Morris Rosenberg, & Leonard I. Pearlin.* Adaptive Behavior in Competent Adolescents: Coping with the Anticipation of College 375
55. *Donald E. Super.* Consistency and Wisdom of Vocational Preference as Indices of Vocational Maturity in the Ninth Grade 385

Intelligence and Creativity

56. *Philip W. Jackson, Jacob W. Getzels, & George A. Xydis.* Psychological Health and Cognitive Functioning in Adolescence: A Multivariate Analysis 392
57. *Ellen V. Piers, Jacqueline M. Daniels, & John F. Quackenbush.* The Identification of Creativity in Adolescents 398
58. *Katherine P. Bradway & Clare W. Thompson.* Intelligence at Adulthood: A Twenty-Five Year Follow-Up 403

Concept Development

59. *Ronald C. Johnson.* A Study of Children's Moral Judgments 413
60. *Bert Y. Kersh.* The Motivating Effect of Learning by Directed Discovery 431
61. *Per Saugstad & Kjell Raaheim.* Problem-Solving, Past Experience, and Availability of Functions 435

Introduction

HOW TO READ THIS BOOK

This book of readings has three main purposes. The first is to place in the hands of the student primary source materials to supplement the textbook. The usual text cannot devote sufficient space to a lengthy consideration of any particular research project or of any particular theoretical position that may emerge from research.

But simply replacing scattered library resources with a collection of readings between two covers is not sufficient. In order to comprehend the structure of the discipline of developmental psychology, students today require an increased understanding of how the developmental psychologist goes about his business. Accordingly, the second purpose of this book is to provide students with examples of the variety of approaches utilized by researchers in their attempts to discover how children grow and develop psychologically. An outline of the research approaches represented in the articles in this volume is given in Table 1, page xv, and a brief description of each research approach is given in the first section of this introduction.

The third purpose, intimately connected with the second, is to provide students with a beginning appreciation of the statistical tools used by the researcher in assessing and interpreting his data. One cannot assume that the student taking a beginning or even a second course in child development, human development, or developmental psychology has any clear understanding of statistics. His exposure to statistics in general psychology is usually brief and not in the context of research literature. The second section of this introduction undertakes to give him a rudimentary acquaintance with these basic statistical tools—enough to understand their use in the various articles in this volume and even to compare the use of different statistical techniques in different situations. It makes no attempt to explain the computations involved but explains only what they establish—what questions they answer and what conclusions they allow the researcher to draw.

Except in Part I, the articles in this collection are organized in an epochal arrangement—that is, according to the age of the child. This is a time-honored approach based on the notion that response, or behavior, is related to age. This ordering, however, should not be taken to mean that age is the most significant factor in behavior or even that it is necessarily a causative factor in behavior. Age is a useful guide, however, because knowing the age of a child generally permits some consideration of the effects of transactions between biological and experiential factors in development.

TYPES OF RESEARCH APPROACHES

Five main types of research approaches are represented in the articles in this collection. These approaches and the papers exemplifying them are listed in Table 1. Whether the articles are assigned by the teacher according to the age sequence in which they appear or according to the research approach represented, students should keep this framework of research approaches in mind while reading and interpreting a given study.

Reviews of the Literature

The first type of research approach consists of articles which are essentially *reviews of the literature* and either are based on a particular theory or a portion of a theory, such as Piaget (6)[1] and Bruner (8), or lead to a theoretical position, such as Hunt (1) and White (2). Such articles do not merely review the empirical evidence but organize this evidence in some systematic, meaningful fashion. Although such a review is not technically a type of research, it represents a useful and important step in the organization and communication of scientific knowledge.

The reviews here deal essentially with such aspects of the child as individuality, consistency, motivation, and cognitive development, and with the techniques used in the assessment of these aspects. The discussion transcends any particular age level and deals either with development over a span of years or with fundamental theoretical constructs. These readings were selected to provide background for understanding the newer ideas about motivation, change and stability through time, and interindividual differences.

Descriptive Studies

The next type of study can be labeled *descriptive*. Such an investigation is not a study in which variables are manipulated in order to influence the outcome but rather a field study of what seems to exist in relatively unselected samples of children and youth. The first major group of these reports focuses on the research done at two centers: namely, the California Growth Center at Berkeley and the Fels Research Center at Antioch, Ohio. These are *longitudinal* studies: measurements of many variables were taken on small samples of youngsters at intervals over a period of years. As Kessen points out, "The longitudinal design is to be favored in the study of those age-functional relationships in which relatively small but individually stable changes over age are expected" (1960, p. 41).[2] The Bayley article (4) is one example of a longitudinal study of age-functional relationships.

There are, of course, problems in the use of the longitudinal approach. Sometimes it is difficult to take into account the effects of cultural variables. Further, it is often difficult to know how far one can legitimately generalize to groups other than the sample being assessed. On the other hand, the longitudinal study offers a decided advantage to the investigator who wishes to examine not only the influence of age but also the influence of life experiences in the development of a child. He can examine the sequence of particular experiences in the life of the child in order to understand their effects at the time and later.

Longitudinal studies do not necessarily originate from a particular theoretical position. A research group sets out to gather data on as many variables as possible that might influence, explain, or demonstrate the development of children. Later on, the group sorts these data to gain understanding and insight. This "atheoretical" approach can yield much information, but sometimes late in the study hypotheses occur to the research group which their data are inadequate to test. Later studies can then be developed to test these hypotheses.

The second major descriptive method is the *cross-sectional* approach, in which groups of children of different ages are studied. This approach has the advantage of time in that one does not have to wait fifteen or twenty years for certain insights. Another advantage is that populations from different subcultures can be utilized so that one can study not only the effect of age but also the effects of different life experiences. One difficulty of this cross-sectional approach is that it must assume that the culture is relatively stable, that age is a more crucial variable than cultural experience within even a given subculture. Thus, if one had studied both six- and fourteen-year-olds in 1948, he would have had to assume that by 1956 no major changes in culture would have occurred to modify the lives of the six-year-olds and that at fourteen they would be comparable to his original group of fourteen-year-olds. But in so doing, he would have overlooked the impact of television and the Korean War. To use the semanticists' symbols, can one assume that fourteen-year-old$_{1948}$ = fourteen-year-old$_{1956}$? Further, the cross-sectional approach cannot deal with the problem dealt with successfully by longitudinal study; the rate of change of behavior in cross-sectional studies is only in static pictures, and each child appears but once. Nor can it show the effect of a particular experience on an individual child. It *can,* however, indicate the effects of a variety of variables—including age—upon behavior if the differences between subjects are large and discrete.

---

[1] The numbers in parentheses throughout this introduction refer to the number of the article in this volume.

[2] The year of publication is given when a study not included in this volume is cited. The full reference is given at the end of this introduction.

This type of design, in which one variable is assessed for its relationship to another, or in which several variables measured at approximately the same time are examined for their interrelationships, has been labeled an *R-R* type. In such a study, responses are related without the manipulation of stimuli, and conclusions specify the degree of relationship. Thus they are essentially correlation studies. It is not possible, from these studies, to talk in terms of antecedent-

---

Table 1: *Types of Research Approaches*

Reviews of the literature

Bruner, J. **(8)**
Caldwell, B. **(13)**
Hunt, J. McV. **(1)**
Kessen, W. **(10)**
Lynn, D. **(9)**
Newton, N. **(16)**
Piaget, J. **(6)**
White, R. **(2)**
King, W. H. **(39)**
Vernon, M. D. **(23)**

Descriptive studies

Longitudinal studies

Bayley, N. **(4)**
Bradway, K., & Thompson, C. **(58)**
Crandall, V., & Preston, A. **(29)**
Jones, M. C., & Mussen, P. H. **(48)**
Lamy, M. **(36)**
Moss, H., & Kagan, J. **(5)**
Mussen, P. H., & Jones, M. C. **(47)**
Tuma, E., & Livson, N. **(50)**

Cross-sectional studies, child's age as the major independent variable

Broderick, C., & Fowler, S. **(38)**
Coleman, J. S. **(51)**
Elkind, D. **(7)**
Johnson, R. C. **(59)**
Kessen, W., *et al.* **(11)**
Stith, M., & Connor, R. **(26)**
Yarrow, M. R., & Campbell, J. **(37)**

Explanatory studies

Parental factors as major independent variables

Crandall, V., & Preston, A. **(29)**
Hoffman, L. **(28)**
Kell, L., & Aldous, J. **(17)**
Klatskin, E., *et al.* **(20)**
Mischel, W. **(30)**
Shaw, M., & Dutton, B. **(49)**
Tuma, E., & Livson, N. **(50)**

(Continued next column)

School situation as major independent variable

Gordon, I. J., & Wood, P. **(32)**
Kitano, H. **(22)**
Kounin, J., & Gump, P. **(31)**

Cultural factors as major independent variables

Anderson, H. H., & Anderson, G. L. **(45)**
Boehm, L. **(40)**
Mischel, W. **(30)**
Schramm, W., *et al.* **(46)**

Child's personality factors as major independent variables

Clarke, H. H., & Clarke, D. H. **(43)**
Gordon, I. J., & Wood, P. **(32)**
Jackson, P. W., *et al.* **(56)**
Jones, M. C., & Mussen, P. H. **(48)**
Kipnis, D. **(53)**
Lamy, M. **(36)**
Mussen, P. H., & Jones, M. C. **(47)**
Muuss, R. **(44)**
Reese, H. **(52)**
Silber, E., *et al.* **(54)**
Super, D. E. **(55)**

Experiments

Crabtree, C. **(35)**
Harlow, H. **(18, 19)**
Kersh, B. **(60)**
Meyerowitz, J. **(27)**
Ojemann, R., & Pritchett, K. **(24)**
Ojemann, R., & Snider, B. **(33)**
Osler, S., & Weiss, S. **(41)**
Saugstad, P., & Raaheim, K. **(61)**
Torrance, P., & Myers, R. **(34)**
Wohlwill, J., & Lowe, R. **(25)**

Development of assessment procedures

Guilford, J. P. **(3)**
Johnson, R. C. **(59)**
Kessen, W., *et al.* **(11)**
Knobloch, H., & Pasamanick, B. **(12)**
Pease, D., & Hawkes, G. **(21)**
Piers, E. V., *et al.* **(57)**
Rheingold, H. **(14)**
Schaefer, E., & Bell, R. **(15)**
Sigel, I. **(42)**

consequent factors, and the reader must be careful not to assume that correlation represents causation.

Explanatory Studies

Although falling within the descriptive category, *explanatory studies* are designed not only to describe relationships but also, because of the variables chosen, to explain behavior. The researcher chooses his variables, usually upon some theoretical basis, and then conducts what might be called a *field investigation*. He selects a sample of subjects and investigates their behavior by seeing if it is related to a particular variable or variables. In the cross-sectional method described above, the main variable with which behavior is related is age. In the explanatory studies described below, age may also be a variable, but, for purposes of simplification, some other factor, such as a cultural condition, is listed as the major independent variable, whose influence on behavior is to be ascertained.

Because it is clearly understood that behavior is not the function of any single variable, most explanatory studies investigate the interactive effects of several variables upon behavior. A good example of this type is Mischel's study (30) in which both the absence of the father and the type of subculture are related to child behavior on a specific task. Nevertheless, for sharper focus, the following discussion is organized around the major independent variable.

The first independent variable listed in Table 1 is *parental factors*.[3] The fundamental question here is how much a parent influences the development of the child. An outstanding example of a study designed to answer this question is that of Sears, Maccoby, and Levin (1957). By means of parental interviews and observation of preschool children, they found relationships between child-rearing practices and child behavior. The validity of parents' self-reports, however, has been seriously questioned by many researchers. One might ask, "Is the behavior of a mother related, and in what way, to her self-report?" The Crandall and Preston study (29) illustrates one approach to answering this question.

The second major variable is the *school situation*. Are such pupil variables as self-esteem related to the behavior of the teacher? Common sense would suggest that they are, but careful study of the pupil-teacher relationship is important in checking on common-sense ideas, in quantifying the degree to which the hypothesis that pupil behavior reflects teacher behavior may be true, and in gaining further insight into the manner in which the relationship may operate. It is equally common sense to assume that parental behavior influences the child, but the changes in advice given to parents by "experts" in the past two decades indicate we are still a long way from a scientific understanding of such adult-child relationships. Only as research yields clues can teacher education and parent education in this area of interpersonal relationships become more scientific and less vague, opinionated, and amorphous.

The Kounin and Gump article (31) is an example of research into the effects of the school situation using a *natural experiment* design. The researchers located teachers who behaved differently toward children and then measured the extent to which the children responded differently to questions about discipline.

Another obvious concept is that children in *different cultures* develop different attitudes and patterns of behavior. This third major variable is still in need of study to delineate the specifics underlying such a broad construct. Two articles here are representative of current investigations. Both the Schramm study (46) of the impact of television and the Andersons' study (45) of cross-national differences in authoritarianism began with hypotheses and studied youngsters who have had different cultural experiences. Both studies found differences among children which were attributable to the cultural variable being studied. Neither research team, of course, would assume that the aspect of culture they investigated is the *only* contributor to cultural differences. These two studies resemble experiments, except that the variable exists naturally and is not manipulated by the researcher. Both are clearly different from the Crandall and Preston study (29), cited earlier, in that they utilize control groups or several populations rather than the single-group pattern of longitudinal research.

The fourth variable is *the child himself*. Most psychologists accept the concept that behavior is a function of the transactions between the person and the environment. A study simply of input (the first three variables mentioned and age) is insufficient, just as a study of a child's behaviors without observation of the setting and situation in which they occur would also be incomplete. How does the child's own personality influence his perception of and his behavior in the situation? And further, how does the child's behavior influence the behavior of those around him? Wenar and Wenar state:

---

[3] Although there actually may be interaction, the researchers whose articles are presented in this group of explanatory studies generally consider the behavior of the parent or the teacher as independent of the behavior of the child. Those investigating the relation of physical maturity to self-concept (Clarke, Mussen and Jones) generally consider the variable of physical growth as not under the influence of the child's personality. An independent variable is that variable which the researcher, in a particular investigation, considers to be uninfluenced by but which itself influences the other variables being measured.

> The matter of child-engendered adult behavior only rarely receives serious attention in empirical studies . . . we are proposing what might be called the "tri-potential" child—one who is now molded by the environment, who now remains unchanged in the face of environmental pressures to the contrary, and who now determines the reactions of the significant people around him (1963, pp. 704–705).

The Gordon and Wood research (32) investigates the latter question in terms of the relationship between the child's estimate of self and the teacher's estimate of him.

How do different aspects of the child's personality relate to each other? How do aspects of the child's personality organization change with age? The studies in this final group of explanatory studies are examples of efforts to answer this type of question. The Clarkes' study (43) deals cross-sectionally with the same problem dealt with in a longitudinal fashion by Mussen and Jones (47, 48): the nature of the relationship between physical body variables and self-assessment.

Several studies utilize measures of self-estimate, usually of the self-report type. For example, Reese (52) and Kipnis (53) both investigate the relationship between self-estimate and social relationships. In both cases, the authors use the term *self-concept* as synonymous with *self-report*. Lamy (36) infers self-perception from behavior, and Mussen and Jones (47, 48) infer self-concept from protective technique scores.

The reader needs to be aware that especially in the area of personality measurement the same term is often defined in operationally different ways. Thus any experimental results must be understood in terms of the theoretical position and the operational definitions of the particular researcher. Of course the same problem exists when one attempts to define concepts in "learning," and the naïve reader is often confused. There are different theories and definitions of learning, just as there are differences in ideas about personality. A simple rule of thumb is to check the usage in the particular research rather than to assume the meaning of any given term.

Experiments

The fourth category of research approaches is *experiments*. These studies are characterized by the careful manipulation of particular identifiable variables, the measurement of the effects of manipulation upon children's behavior, and the use of a control group. The utilization of a control population is particularly important in studies of children because of the ever-present question of the influence of growth. At times a control group is not part of the design, but control results from knowledge obtained previously from descriptive studies. For example, Harlow (18, 19) did not utilize controls, but certainly the behavior of the monkeys described in his second article is atypical and can be attributed to the treatment they received. If children were measured on an achievement test at the beginning and end of a training situation which lasted for a semester, however, a control group would be necessary to determine whether it was the training, the passage of time, or some other variables which brought about any changes in their performance on the second test.

If a researcher is interested in the question of whether children of a given age or ability can learn a particular skill or concept, then a control population may not be crucial. Torrance and Meyers (34) demonstrate that bright elementary-school youngsters can learn certain research skills, and Ojemann and Pritchett (24) show that through a guided learning experience kindergarten children can comprehend *specific gravity*. In the latter case, as with Harlow's monkeys, a sort of control exists in the natural population. The careful reader must evaluate the purposes of a study and the generalizations drawn by the author in order to determine whether a control group of some sort would be necessary if one wanted to extend the results beyond the group studied.

A number of the experiments included here compare several types of training. For example, Wohlwill and Lowe (25) use three training methods and a control group, and Kersh (60) uses three teaching techniques to assess the relative superiority of one method over another. Wohlwill and Lowe are concerned with testing Piaget's theory (6), and Kersh with testing Bruner's (8). Their various procedures for training or teaching are related to these respective theories. This is a more powerful approach than the strictly empirical one of proceeding in a trial-and-error fashion without guidance from theory.

The article by Wohlwill and Lowe illustrates the flow of research from one procedure and level of operation to another. First, we have Piaget's theory of cognitive development, based upon cross-sectional descriptive work. Next, we have Wohlwill and Lowe using experimental methods to test an aspect of the theory in which they use previous data from descriptive studies by Piaget as a basis for selecting the age group to be used as subjects. Finally, we have their recommendation for a longitudinal study to test the theory further. In this way is our knowledge built.

Development of Assessment Procedures

The last research approach represented in the articles in this book is that of *assessment procedures*. These procedures are crucial for the obvious reason that even when all other aspects of research design are excellent, the findings are still only as good as the instruments utilized in collecting the data.

Essentially, the developmental psychologist uses either direct or indirect observation of the person. The choice between these may be determined by theoretical predictions, by whether he believes overt behavior conceals more than it reveals, or by practical considerations. The use of self-reports by mothers and children is a case in point. Although observation in the home may be desirable, an interview or rating schedule may be much cheaper and almost as good. Its validity must be tested, however, before it can become a legitimate substitute. Caldwell (13) indicates some of the problems involved in assessment of infant personality. The Parental Attitude Research Instrument (PARI), developed by Schaefer and Bell (15), is an example of an instrument designed to measure indirectly what it is difficult to observe.

Improvement in assessment can come from a variety of sources. Kessen and his colleagues (11) illustrate how technology can help. Psychological measurement inevitably rests on theoretical bases, for it is always constructs we are measuring. Thus, as a conceptual model of the child is developed, or as our constructs of psychological variables like motivation and intelligence are reorganized and refined, new suggestions for assessment emerge. For example, Knobloch and Pasamanick (12) demonstrate how a conceptual model of the infant organism influences clinical assessment. The discussions by Guilford (3) and Sigel (42) require changes in the concepts of the nature of intelligence from those held by many psychologists and educators in the past decades. How intelligence—and personality—is measured is related to how it is defined.

We see, then, that although students of human development use a variety of research designs and measurement procedures and represent different theoretical positions, their central focus is the same: the growing child's behavior and development.

STATISTICAL TECHNIQUES

In addition to developing instruments, constructing theories, and selecting research problems, the investigator must describe the results of his experiments and draw conclusions from them. To describe economically the behavior of the subjects in any study, *descriptive statistics* are employed. To assess the probability that similar individuals in similar circumstances will act the way the subjects did, the investigator makes use of *inferential statistics*. Description and inference are basic to science, so it is not surprising to find both techniques used in the papers in this volume. It should also be apparent that comprehension of research reports will depend on familiarity with these statistical techniques.

This section is intended to provide a knowledge of statistics adequate for an understanding of the readings in this book. It does not attempt to deal with all the statistical procedures used by developmental psychologists in their work or with the complex issues involved in both the choice and the utilization of these procedures. This wider background, and the greater critical understanding that it implies, must be left to courses in statistics and research methodology.

Table 2, on page xix, arranges the readings according to the type of statistics they employ. The student may wish to test his understanding of a given concept by examining the articles listed under that heading in the table. Some descriptive techniques, such as the mean and the standard deviation, are quite extensively used and therefore have not been tabulated.

Descriptive Statistics

In general, the results of a research study are numbers. These numbers are the raw data, the measures of the way individual subjects performed on some test or task. Descriptive statistics are applied to the raw data to organize it meaningfully or to summarize it efficiently. Results could be reported simply by listing the scores made by all the subjects, but it would be difficult to scan such a mass of numbers and gain an accurate impression of how the group performed as a whole. And if hundreds of individuals are tested, it is not economical to report all the scores.

The median. If we were to choose one score that best represents a group as a whole, we might pick the *median*. The median is the score which divides a group into two equal-sized subgroups. Half the scores are above the median, and half are below it. In a sense, then, the median is the middle score, and in this sense it represents the entire group of scores.

The mean. The median simply divides the scores into two equal halves; it does not take into account how far the highest and lowest scores are from the median. Suppose the median score in a group of nine scores is 100, but four people got 99 while the remaining four got 150. The median would not represent this group very well, because the size of the scores is ignored. The magnitude of each score is taken into account if we add up all the scores in the group and divide by the number of scores. The result is what is often called an average; in statistics it is called the *mean*.

The range. The mean and the median both provide an average value that indicates the performance of the group as a whole. But they do not tell us whether most members of the group were "about average" or whether the individual scores varied widely. The simplest way of indicating the amount of spread, or variability, in the scores of a group is to compare the highest score with the lowest. The difference between these extreme scores is called the *range*. Thus if the highest score in a group is 150 and the lowest is 100, the range is 50.

The variance and the standard deviation. The range is a measure of how closely scores are clustered together. But it is a relatively crude measure, because it is based on only two scores: the highest and the lowest. The *variance* and the *standard deviation* are more refined measures of variability which are based on all the scores in the group. They take into account the deviation of each score from a central reference point—the mean. When every score is close to the mean, the overall variability is slight, and the variance and the standard deviation are small. When

Table 2: *Types of Statistical Analyses*

Correlation

Bradway, K., & Thompson, C. **(58)**
Crandall, V., & Preston, A. **(29)**
Jackson, P. W., *et al.* **(56)**
Johnson, R. C. **(59)**
Kessen, W., *et al.* **(11)**
Kipnis, D. **(53)**
Knobloch, H., & Pasamanick, B. **(12)**
Lamy, M. **(36)**
Moss, H., & Kagan, J. **(5)**
Muuss, R. **(44)**
Osler, S., & Weiss, S. **(41)**
Piers, E. V., *et al.* **(57)**
Rheingold, H. **(14)**
Schaefer, E., & Ball, R. **(15)**
Schramm, W., *et al.* **(46)**
Super, D. E. **(55)**
Tuma, E., & Livson, N. **(50)**

Factor analysis

Crandall, V., & Preston, A. **(29)**
Guilford, J. P. **(3)**
Meyerowitz, J. **(27)**

Chi square

Anderson, H. H., & Anderson, G. L. **(45)**
Boehm, L. **(40)**
Bradway, K., & Thompson, C. **(58)**
Gordon, I. J., & Wood, P. **(32)**
Jackson, P. W., *et al.* **(56)**
Kell, L., & Aldous, J. **(17)**
Kersh, B. **(60)**
Klatskin, E., *et al.* **(20)**
Jones, M. C., & Mussen, P. H. **(48)**
Kounin, J., & Gump, P. **(31)**
Mischel, W. **(30)**
Moss, H., & Kagan, J. **(5)**
Mussen, P. H., & Jones, M. C. **(47)**
Muuss, R. **(44)**
Pease, D., & Hawkes, G. **(21)**
Saugstad, P., & Raaheim, K. **(61)**
Stith, M., & Connor, R. **(26)**
Yarrow, M. R., & Campbell, J. **(37)**

The t test

Clarke, H. H., & Clarke, D. H. **(43)**
Hoffman, L. **(28)**
Kipnis, D. **(53)**
Kitano, H. **(22)**
Ojemann, R., & Pritchett, K. **(24)**
Ojemann, R., & Snider, B. **(33)**
Shaw, M., & Dutton, B. **(49)**

Analysis of variance

Crabtree, C. **(35)**
Elkind, D. **(7)**
Kessen, W., *et al.* **(11)**
Muuss, R. **(44)**
Osler, S., & Weiss, S. **(41)**
Reese, H. **(52)**
Wohlwill, J., & Lowe, R. **(25)**

scores are spread out, the difference between each score and the mean is relatively large, and the variance and standard deviation are correspondingly greater.

Correlation. Psychologists are often concerned with two or more variables and the relationship between them. Let us consider the kinds of relationships we might find between college entrance examination scores and college grade-point averages. First, the variables might be *positively correlated.* If so, individuals with high examination scores will have high grade averages. If the relationship is perfect, an individual's rank on one variable will be the same as his rank on the other. Thus, in the case of a perfect positive correlation, an individual's standing on one variable can be predicted exactly if we know his standing on the other variable. For example, the person who is second best on the examination will have the second highest grade average.

If the variables are *negatively correlated,* rank on one variable is inversely related to rank on the other. Thus an individual with a high examination score would have a low grade average, while an individual with a low examination score would have a high grade average. Notice that if there is a perfect negative correlation, exact prediction is still possible. For example, the person with the third highest examination score will have the third from the lowest grade average.

Now consider the possibility that there is no relationship between the two variables. In this situation, the individual with the highest examination score might have high, low, or average grades, and the same is true of those with low examination scores. When an individual's position on one variable is thus independent of his position on the other variable, we say there is no correlation between the variables.

Extreme cases, in which two variables have a perfect positive or negative correlation, are unusual. But it is not difficult to imagine the more common intermediate situations. Thus a high correlation is one that comes near to being perfect. When a correlation is high, predictability is high but not perfect. An individual with a high rank on the examination *might* have high, low, or average grades, but it is much more probable that his grades will be high. When variables are relatively independent, the correlation between them is relatively low. In such cases, an individual with a high rank on one variable might be either high, low, or average on the other variable, but he is only somewhat more likely to be high.

The amount of correlation between two variables is expressed in terms of a *correlation coefficient.* In a perfect positive relationship, the coefficient of correlation is +1.00, whereas a perfect negative correlation has a coefficient of −1.00. A coefficient of 0.00 indicates no correlation. Intermediate degrees of correlation are represented by coefficients between one and zero. Thus coefficients of .85 and .92 represent high positive correlations; coefficients of −.85 and −.92 represent high negative correlations.

The naïve reader is tempted to assume that correlation coefficients are like percentages, but they are not. A coefficient of .60 is not a 10 per cent better predictor than one of .50. The use of a correlation coefficient below .85 for individual prediction is extremely hazardous.

The researcher is interested in determining relationship and studying its nature. The administrator, clinician, or teacher is interested in application. A relationship may be low, but still existent, without being of practical importance. The student should be aware of the difference between the foci of the researcher and of the applier. The applier, in turn, must understand the need for a high level of relationship before applying research findings to administrative or professional problems.

Two general uses of correlation are found in the studies in this book. Many research designs call for observers to rate behaviors, or traits, or performance. Usually, several judges or observers are used. One use of correlation relates to the measurement of the reliability, or sameness, of the judges' ratings. Thus, the more Judge A's ratings agree with Judge B's, the higher the degree of correlation. Moss and Kagan (5), Knobloch and Pasamanick (12), and Kessen (11) use this procedure.

In the second type of correlation, cited earlier, two variables are related. These variables may be test scores, behavior ratings in two situations, or any other quantifiable activity. The type of quantification used will determine the actual correlation measure applied to the data. Sometimes, because of the nature of the data, or the sample, it is not possible to compute the correlation coefficient (r) by using raw score measures. Thus, there are other coefficients in addition to r.

Rho is a rank-order coefficient in which the subjects are ranked on performance one and performance two, and the test is made to see if their relative rank on one is related to their rank on the other. Phi is used in cases in which data cannot be ordered, or ranked, but in which subjects can be counted. An example of the latter would be a statement of the relationship between marriage and success in teaching for women. The population can be divided into discrete groups—married, single, divorced, widowed—and the proportion of women in each group who are rated as successful and unsuccessful can be calculated to produce the phi coefficient. Although these two coefficients are often viewed as weaker tests of relationships than r, they allow the researcher more flexibility in design. All correlation procedures, however, are essentially designed to elicit the degree of relationship between variables.

Factor analysis. The psychologist often wishes to explore the interrelationships between a large number of variables. For this purpose he may use the technique of *factor analysis.* Factor analysis is based on the correlation of each variable with all the other variables used in the study. When a number of variables all correlate with one another, the existence of a common factor is tentatively assumed.

Suppose, for example, that a group of people is given a battery of tests. The battery includes tests of reading ability, ability to think of and recognize synonyms and antonyms, and ability to define words. If these three tests correlate highly with one another, it is reasonable to assume that there is some factor common to all of them. We might suppose that there is a factor of verbal ability which is being measured to some extent by each of the three tests.

We might find a similar interrelationship between another group of tests dealing with the manipulation of numbers and assume that these tests are all measuring aspects of a mathematical ability factor. And if it is then found that mathematical ability and verbal ability are correlated to some degree, a general intelligence factor might be postulated.

For example, Crandall and Preston (29) used factor analysis in studying maternal behavior and identified three major dimensions: affectional, protective, and coercive control. In each case, several scales of behavior were found to be intercorrelated sufficiently to justify the assumption that each scale was measuring aspects of a common dimension. In the case of intelligence testing, Guilford (3) has done essentially the same thing in order to locate intelligence test scales that are related to each other sufficiently to be assumed they are measuring aspects of the same factor.

It is important to realize that the results of a factor analytic study depend on the variables that have been measured, the measuring instrument, and the sample. If the battery in the example had not included tests of mathematical ability, the investigator might have assumed that general intelligence was the only factor common to the three verbal tests. He would have had no reason to postulate separate verbal and mathematical factors, partially independent of each other. If he had used other tests or tested other groups, different factors might have emerged. Thus, while factor analysis is extremely helpful in identifying relationships, it should be remembered that the interpretation of these relationships requires caution.

Inferential Statistics

The techniques discussed in the previous section are concerned with describing the performance of actual subjects who have participated in a research study. The subjects may be thought of as a *sample* of a larger group. It is this larger group, or *population,* in which the experimenter is interested: he wishes to know whether the behavior of the sample is characteristic of all people, or all college students, or all adult American males, or some other defined population.

To answer this question, the experimenter makes use of *inferential statistics.* The answer is never certain; there is always the possibility that his results are due to chance. However, by using inferential statistics, he can determine the *probability* that only the laws of chance are operating to produce his result.

The probability that a result is significant, rather than due to chance, is reported in a statement such as $P < .05$. This statement may be read, "The probability that the results were due to chance is less than 5 in 100." Since .05 is the same as $^{5}/_{100}$ or $^{1}/_{20}$, the statement means the odds are less than 1 in 20 that the results were due to chance. Because $^{5}/_{100} = 5\%$, the *level of significance* in this case is 5 per cent. At the 1 per cent level of significance, the probability that the result would occur by chance is only 1 in 100; in other words, there is a 99% chance that the results are significant.

It should be realized that statistical significance is not the same thing as "psychological" significance. Statistics can tell us only how likely it is that a result is due to chance. It is the individual researcher who must decide whether the odds against chance are great enough to justify accepting the result as significant. His decision will depend on a number of things. He may also be influenced by the reasonableness of the result. If the outcome of the experiment is consistent with existing facts and theories, he will be inclined to accept it. But if the outcome seems inexplicable, or if it contradicts the results of other experiments, he may withhold his acceptance until further research has confirmed his findings. If he then refines his measures and his research design, he may demand a higher level of significance.

In any case, the researcher should report the probability level he has obtained. With this information, the reader is in a position to draw his own conclusions about the significance of the result. We shall briefly discuss three important procedures for determining statistical significance. All these methods yield probability statements of the kind we have examined.

Chi square. Suppose an investigator wishes to discover whether boys perform differently from girls on a test of imagination. He tests a mixed group and finds the group mean. He then formulates the *null hypothesis:* he tentatively assumes that boys and girls are equal in their performance on his test—that there is *no* difference. If this is true, approximately equal numbers of boys and girls should score above the mean. If 50 boys and 50 girls are tested, the null hypothesis predicts that 25 boys and 25 girls will be above the mean. The investigator may then compare this expected distribution with the distribution he actually found by means of a statistic called *chi square* (symbolized χ^2).

If the obtained distribution is very close to the distribution predicted by the null hypothesis, the assumption of no difference is confirmed. Thus if 26 boys and 24 girls are above the mean, the slight difference between the obtained and the expected distributions would probably be due to chance. On the other hand, suppose 45 girls and 5 boys are above the mean. Chi square would show that the difference between this result and the predicted result is significant. In other words, it is not likely that this distribution would occur if there were not a real difference between boys and girls in regard to this test. Exactly how unlikely this possibility is will be indicated by the probability level obtained by chi-square analysis.

The t test. Experimenters are often interested in the significance of the difference between two means. Suppose an investigator gives a personal adjustment test to both normal and "problem" children. He finds that the mean for the normal children is higher than the mean for the problem children. The question is: does the difference reflect a real personality difference or is it the result of chance? The *t test* is a statistic designed to assess the probability that the difference is due to chance. Other things being equal, the larger the difference between the means, the greater the probability that the difference is significant. On the other hand, the greater the variability within the groups, the more they will overlap and thus the less meaningful is a given difference. The *t* test takes into account both the difference between means and also the variability of the groups in determining the possibility that chance is the cause of the difference.

The size of the sample, as well as its composition, often determines the type of statistical test the researcher will use. Obviously, the larger the sample being studied (if it has been selected correctly), the safer it is to assume that its distribution of scores will be "normal"—that is, falling into the shape of the classic bell-shaped curve. Most college students, particularly those who have been in large classes, have been graded "on the curve." They realize that, for large classes, this is probably as fair as any system, but they usually react negatively when a teacher in a class of less than thirty students marks this way. They recognize, intuitively, that this assumption of normal distribution does not necessarily hold—they see it as limiting them and creating unfair competition. The statistician, too, recognizes that the shape of the curve changes with small populations. The *t* test was designed to be used with samples of less than thirty in order to allow for some of the change in the shape of the bell. However, the assumption is still present that even the small population is normal for its size and does not contain an undue amount of people who score at one end or the other. If this assumption is met, the *t* test can be used.

An example of the use of this test is in the Kitano study (22). Here, teachers rated pupils on a behavior scale based upon their observations. The mean score and standard deviation for each group (problem and normal) were computed for each situation rated, and then the *t* test was used to see if the difference between the means of the groups was statistically significant. This is the usual procedure followed.

The *t* test is one way in which the researcher uses statistical devices to cope with the problem of size of the sample. The student must not assume that a small sample is bad and that a large one is good. The more important criteria are how the sample was selected, how well it represents the total universe of children from which it was drawn, and whether the researcher has used statistics appropriate to its size and composition.

Analysis of variance. Techniques like the *t* test deal only with two groups. The psychologist, however, is often interested in the differences among a number of groups. For example, he may wish to compare the relative efficiency of three verbal training procedures. Thus, he might need to assess the significance of differences among four groups: a control group and three groups trained by different procedures. The Wohlwill and Lowe study (25) is a good example. There were 72 children, divided into 4 groups of 18. Each group was trained differently (4 conditions of training) and measured before and after training (2 times on two tests [verbal and nonverbal conservation]). *Analysis of variance* makes such comparisons possible. This procedure examines the amount of variability within each group and the extent to which the groups differ from one another. The significance of any differences among the groups is stated in terms of probability, just as in the case of the other inferential techniques we have discussed.

CONCLUSION

The articles in this book can be approached and studied in several ways. The sequence of placement follows an age pattern; this introduction has stressed a research orientation. It is possible to use this collection not only to provide research findings but also to teach how research is conducted.

A third categorization of the material transcends both age and research design. This category is the topic. Parts II through V are subdivided by topic, some of which run throughout the book. For example, parent-child relationships can be examined at each age level. Work on cognitive development can be found throughout the book, from Bruner (8) and Piaget (6) in Part I to Saugstad and Raaheim (61) in Part V. It is suggested that these various categories be used in a flexible fashion to obtain the best use of these materials.

REFERENCES

KESSEN, W. Research design in the study of developmental problems. In P. H. Mussen (Ed), *Handbook of research methods in child development.* New York: Wiley, 1960.

SEARS, R. R., MACCOBY, ELEANOR E., & LEVIN, H. *Patterns of child rearing.* Evanston, Ill.: Row, Peterson, 1957.

WENAR, C., & WENAR, S. The short term prospective model, the illusion of time, and the tabula rasa child. *Child Develpm.*, 1963, *34*, 697–708.

PART I

GENERAL CONCEPTS

GENERAL CONCEPTS

1

Experience and the Development of Motivation: Some Reinterpretations

J. McVicker Hunt

Hunt's current interest centers on the changing beliefs about fixed intelligence and predetermined development. This interest is reflected in his recent book, Intelligence and Experience *(1961), in his papers on stimulating intellectual development in preschool children, and in this article. Here, the author reviews the data about motivation, stressing the current movement away from a mechanistic, drive-reduction position.*

A recent issue of the *Saturday Evening Post* carried a cartoon that some of you may have noted. It depicts a boy entering his house, perhaps from school, where his father is sitting with his paper. The boy appears to be fixing his father with an accusing glare. The punch-line reads, "Somebody goofed. I'm improperly motivated."

This cartoon depicts the vantage point from which I have been examining what we think we know about the relation between experience and motivation. When a child's behavior fails to fit the standards somebody in our society holds for him, it is pretty well agreed among us who are supposed to be experts on human nature that "somebody goofed." And that somebody is usually considered to be a parent.

The question is: what is the proper formula? If one examines the accruing evidence relevant to what has been the dominant conception of the experiential sources of motivation, one can hardly escape the conclusion that this conceptual scheme needs some revisions. If we based our child-rearing entirely on our dominant theory of motivational development, we would probably goof as often and as badly as run-of-the-mill parents.

Today I wish, first, to remind you of three of the most basic and general of the propositions in that theory of motivation which has been dominant for the past 30 to 40 years. These are propositions which, although stated in somewhat varied forms, have been shared by both psychoanalysts and academic behavior theorists. Secondly, I wish to cite evidence which calls these propositions into question, and thirdly, to suggest tentatively three new interpretative principles which appear to me to be congruent with a large number of facts and which have interesting implications.

Our conceptions of motivation have traditionally been concerned with three large questions. (a) Why does an organism or person become active? (b) Why does the organism or person act one way rather than another? and (c) How do you get the organism or person to change his

---

From *Child Development*, 1960, *31*: 489–504. Reprinted with permission of the Society for Research in Child Development.

behavior to something conceived to be more desirable or appropriate?

THE DOMINANT THEORY

Drive

According to our dominant theory, it is claimed, first of all, that "all behavior is motivated," and that the aim or function of every instinct, defense, action, or habit is to reduce or eliminate stimulation or excitation within the nervous system. It is not easy to state when this view was first presented. . . .

It is assumed, as the proposition that "all behavior is motivated" implies, that the organism would be inactive unless driven by either inner or outer stimuli. Freud (1950) has been highly explicit about this assumption, and the assumption lies implicitly behind the notion of conditioned or learned drive in behavior theory and behind the traumatic notion of anxiety in psychoanalysis. It is sometimes obvious, of course, that animals and people are sometimes active when it is hard to see how either homeostatic drive or painful external stimulation could be operative. It is then assumed that some of the weak, innocuous stimuli present must have been associated in the past with either painful stimuli or homeostatic needs. In such a way the weak stimuli which are present must have acquired the capacity to arouse the drive, often now called anxiety by psychologists as well as psychoanalysts, and it is such acquired or conditioned drive that is conceived to activate the organism.

Such conditioned drive or anxiety has been well demonstrated in the laboratory. . . . In my own work (1941) rats fed irregularly in infancy ate more and sometimes (1947) hoarded more than their litter-mate controls in adulthood after a period without food. Here, as I conceived it, the cues of hunger were conditioned to intense hunger excitement during the infantile experience. In adulthood the conditioned hunger drive facilitated the rate of eating and, sometimes, hoarding.

Such work has demonstrated that this notion of conditioned drive or anxiety . . . has a solid basis in reality. But in what has been the dominant theory of motivation, as epitomized by Freud's (1936) later traumatic theory of anxiety and by the Hull (1943) and Dollard-Miller (1950, 1941) theory of acquired drives, conditioning is conceived to be the only way in which an organism can become fearful of innocuous stimuli.

Habit

Habit has been the answer to the second question concerned with why an animal or person acts one way rather than another. The organism is controlled by the habits which have served to reduce drive in the past when that organism was in the presence of the inner and outer drive stimuli and the cue stimuli impinging upon him at any given now. Under the term *habit,* I am including psychoanalytic modes, which have supposedly been fixated during infancy in the course of either too much gratification or too much frustration, and I am including also ego-defenses, or anxiety equivalents, and cathexes, as well as the instrumental responses and traits commonly investigated in psychological laboratories.

Changing behavior has been conceived to be a matter of motivating the organism with either punishment or homeostatic need to make the desired behavior which can then be reinforced by arranging for it to reduce the drive aroused by the punishment or the need. Although the conditions and conceptions of psychotherapy in the clinic differ considerably from the conditions and conceptions of the behavior theorist investigating learning in laboratory animals, in either case it is conceived that motivation is a necessity, and motivation means changing the emotional or drive conditions which are quite extrinsic to either the instrumental behavior or the cognitive, informational processes concerned.

This dominant theory has been a conceptual edifice of large dimensions and of considerable detail. It has provided a plausible account of both personality development and social motives. The experimental facts of homeostasis and of conditioned drive and fear are sound. Nevertheless, it has become more and more evident in the past 10 years that some of the basic assumptions of this dominant theoretical scheme and some of the explanatory extrapolations contradict facts and call for reinterpretation.

REINTERPRETATIONS

Is All Behavior Motivated?

The first of the assumptions to be called into question is the one that *all behavior is motivated* and that *organisms become inactive unless stimulated* by homeostatic need or painful stimulation or conditional stimuli for these. A large variety of observations contradict this assumption and imply spontaneous molar activity. Beach (1945) has reviewed the observations of play in the young to show that playful activities are most likely to occur when either young animals or children are homeostatically satisfied and also comfortably warm. The very occurrence of either homeostatic need or strong external stimulation stops play and turns the young animal or child to activities calculated to relieve such stimulation. . . . Harlow, Harlow, and Meyer (1950) have demonstrated that well-fed monkeys will learn to unassemble a three-device puzzle with no other drive and "no other reward than the privilege of unassembling it." In another study Harlow (1950) found two well-fed and well-watered monkeys worked repeatedly at unassembling a six-device puzzle for 10 continuous hours, and they were still showing what he characterized as enthusiasm for their work on the tenth hour of testing. From his observations of the human child, moreover, Piaget (1952) remarks repeatedly on the enthusiastic and repeated performance of such emerging skills as the release of a toy, sitting up, standing, etc.

Such evidences of spontaneous behavior, which is unmotivated in the traditional sense, have led to the naming of such new motives as a curiosity drive by Berlyne, an exploratory drive by Montgomery, and exteroceptive and curiosity drives by Harlow. I would like to object that merely naming such drives explains nothing. If we continue, we shall be revisiting McDougall's practice of postulating a separate drive for almost every variety of activities. Let us stop with noting that such observations do contradict our assumption that organisms will become inactive unless driven by homeostatic needs and painful stimuli and give up this ancient Greek notion that living matter is inert substance to which motion must be imparted by extrinsic forces. We can then embrace the thermodynamic conception of living things as open systems of energy exchange which exhibit activity intrinsically and upon which stimuli have a modulating effect, but not an initiating effect.

This notion of activity being intrinsic in living tissue is receiving support from studies of organ systems as well as studies of molar organisms. The EEG, for example, shows that brain cells are continuously active (Jasper, 1937; Prosser, 1934). In sleep the slow waves of large amplitude are taken to imply that large numbers of cells are firing synchronously, and the effect of waking and stimulation and exciting the brain-stem-reticular formation is to asynchronize this firing which shows in rapid waves of low magnitude (Lindsley, 1957).

Granit (1955) points out that the spontaneous firing of retinal cells increases with dark adaptation and thereby functions to prevent the deafferentization of visual cortex with darkness. Twenty years ago, this spontaneous firing was considered, at worst, to be due to some failure of experimental control, or at best, noise in the channel of information. Recently, the Laceys (1958) have found spontaneous fluctuations of sudomotor activity and cardiac activity which they also see as functioning in the control of the organism's relations with its environment. Especially intriguing is their notion that the carotid sinus mechanism functions as a feedback loop which participates in the directing of attention inward or outward by inhibiting or facilitating receptor inputs. But the point of mentioning these evidences of spontaneous activities of organ systems here is merely to help inter for good the notion that activity of living systems requires homeostatic need or painful external stimulation and to foster the idea that to live means to be active in some degree.

Reinforcement

This idea of activity being intrinsic in living organisms has implications for our conception of reinforcement. It makes it unnecessary to see all activity as a matter of either reducing or avoiding stimulation which is implied in the assumption that organisms become inactive unless stimulated. This is a second fundamental assumption of the dominant theory which has been shared by psychoanalysts and behavior theorists alike.

On the one hand, there is still a place for drive reduction. It is clear that under conditions of homeostatic need and painful stimulation, and perhaps under circumstances when the conditions of stimulation are changing with too great rapidity, both animals and persons learn techniques and strategies leading to gratification or reduction in external stimulation. The evidence that led Thorndike to formulate the "law of effect" is as convincing as ever. Moreover, in association with reductions of homeostatic need, animals and men may also learn cathexes or emotional attachments. The facts referred to are those highly familiar in secondary reinforcement. . . .

Especially important in this connection are the studies of human behavior under conditions of minimal variation in stimulation. I refer to the studies of perceptual isolation by Bexton, Heron, and Scott (1954) at McGill and also the work of Lilly (1956). At McGill, college students were paid 20 dollars a day to do nothing. They lay for 24 hours a day on a comfortable bed. The temperature was optimal and constant. Eyes, ears, and hands were shielded to minimize stimulus variation. Few subjects could endure more than two or three days of such conditions. They developed a desire for variation which was almost overwhelming.

While interpreting such facts in terms of a multiple set of drives for curiosity, exploration, or stimulation will get us only to a redescription of them, Hebb's (1955) notion of an optimal level of activation—and, I would like to add, stimulus variation below which *increases* are reinforcing and above which *decreases* are reinforcing—is an integrative conception of fair magnitude. Moreover, the drive-reduction principle of reinforcement may be seen to be but half of this more general curvilinear principle.

But this is probably not the whole story. It looks as if there were natively both positive and negative forms of exciting stimulation. . . . From the fact that Davis (1957) and his collaborators at Indiana have shown that showing pictures of nude women to college males increases excitement as shown by increased palmar conductance and the arrest of EEG-alpha, it is clear that such stimulation is exciting rather than excitement-reducing. . . .

The suggestion that the two forms of excitation, one positive and one negative, are built into organisms comes also from the work of Olds and Milner (1954). Electrical stimulation of the septal area is positively reinforcing, but electrical stimulation of the brain-stem reticular formation is negatively reinforcing. Perhaps, it is not without significance that the septal area is part of the old olfactory brain which has been considered to have an especially important part in the mediation of sexual and consummatory behavior in mammals. At any rate, it looks as though certain types of stimulation may be positively reinforcing even though they be intense and exciting. This may mean that the curvilinear principle may be limited in its domain to strong stimulation via the exteroceptors when homeostatic needs are minimized.

The suggestion of innate, positive, and negative exteroceptive stimulation comes secondly from recent work by Harlow (1958). It has been customary to see an infant's cathexis or love for its mother developing as secondary reinforcement largely out of its feeding experiences. Freud (1938), of course, contended that the pleasure from stimulation of the oral erogenous zone furnished the experiential basis for both pleasure-sucking and maternal attachment, a contention which contradicted his most definitive formulations of drive theory (1950). The fact that an infant must suck for its nourishment, according to libido theory (1938, p. 587), merely guaranteed discovery of the pleasures of oral stimulation. Behavior theorists have seen both sucking and love of mother as forms of secondary reinforcement deriving from the fact that the child satisfies its hunger by means of sucking the mother's breasts (Mussen and Conger, 1956, pp. 137 ff.). . . . Harlow argues from these findings that it is the sensory quality of softness which gives the reinforcement. His study suggests, moreover, that it is important to investigate the capacity for various kinds of stimuli for positive and negative reinforcement in the very young. . . . In view of the finding by Davis (1957) and his collaborators that seeing a picture of a nude female results in reduction in the heart rate of male college students, it is possible that this physiological indicator may provide a technique for determining the direction of the reinforcing effect of stimuli in the newborn. . . .

Conditioned Fear and Anxiety

The third assumption that I wish to examine in the light of empirical evidence is the notion that fear and anxiety are *always* inculcated as a consequence of traumatic experiences of helplessness in the face of homeostatic need or painful external stimulation. Note that I am not denying that such conditioned fears do exist. I am only questioning the word *always* . . . are always inculcated as a consequence of traumatic experiences.

The first relevant studies go way back to the 1920's. Harold and Mary Cover Jones (1928) attempted to test the claims of Watson (1928) and Watson and Raynor (1920) concerning conditioned fears. They exposed their subjects of various ages, ranging from early infancy to adult, to a large but sluggish and harmless bull-snake. Fear of the snake was exceedingly common among adults, teenagers, and latency-age children, but it was absent in children below three years of age. It began to appear among children older than three and was typical of children six and older. From the fact that the fear appeared at a younger age in those of higher intelligence than those of lower intelligence, the Joneses argued that fear of snakes is a response which comes automatically into the developing child's repertoire through maturation. This remains as an alternative hypothesis to that of conditioned fear.

A study by Frances Holmes (1935), which is seldom cited, calls both of these interpretations into question. Holmes compared the fearfulness of the children of lower-class background, who were attending a day nursery, with the fearfulness of children of upper-class background, who were attending a private nursery school. She got her fear scores by indicating that the child could get some attractive toys with which to play by going into the dark room adjacent to the examining room, or by taking them off a chair situated beside that of a strange woman dressed in a large floppy black hat and a long gray coat, or by climbing along a plank some three feet off the floor. If the child started immediately for the toys, he got a score of one for that item. If he hesitated but ultimately went ahead on his own, he got a score of two. If he would go only if accompanied by the examiner, the score was three. If he refused to go at all, the score was four. There were seven such situations. The results show that the fear scores of the lower-class children averaged only about half the size of those for the upper-class children, and the fear scores for boys were lower than those for girls. Yet it would be the lower-class children who had experienced the more homeostatic need and painfully rough treatment than the upper-class children, and the boys had probably experienced more painful experiences than the little girls. That intelligence is not the factor is shown by the fact that the fear scores showed a correlation of only about +.2 with mental age, and the differences were still significant when intelligence was partialed out. Something besides either conditioned fear or the correlation between fear and intelligence is required to make these results comprehensible.

Recently evidence even more contradictory to the notion of conditioned fears has been coming from the work of Seymour Levine. Levine, Chevalier, and Korchin (1956) have compared the adult behavior of rats shocked and rats petted daily from birth to their 20th day with the adult behavior of rats left continuously in their nests with their mothers. When he started this work, Levine expected to find that the shocked animals would show traumatic effects of their shock experiences in heightened emotionality and damaged capacity to learn adaptive responses. On the contrary, the shocked animals, along with the handled animals, gained weight faster than those left in the nest. Byron Lindholm, working with the writer, has repeated and confirmed this finding. Moreover, Levine's shocked and handled animals both showed less emotionality than those left continuously in the nest with their mothers, i.e., less emotionality in the sense that they defecated and urinated less frequently when placed in a strange situation. Finally, the shocked and handled animals, which have appeared alike in all of these experiments, learned an avoidance response more rapidly and drank more readily after 18 hours without water than did the rats left in the nest with their mother.

Clearly these results on both human children and rats imply that fear and anxiety must sometimes have some other basis than that of being associated with painful stimulation. As many of you know, Hebb (1946, 1949) has formulated a radically different explanation of fear which may be termed either an incongruity or a dissonance theory.

The facts which suggested Hebb's conception came largely from observing chimpanzees being raised under controlled conditions at the Yerkes Laboratory. Fear, defined as withdrawal behavior in response to the appearance of some object, does not appear in young chimpanzees until they are approximately four months old. Then, the objects feared are familiar objects in unfamiliar guise. Fear of strangers is an example. This appears spontaneously to the first stranger seen, so it cannot be based on associating strangers with painful stimulation. Fear of strangers does not appear in chimpanzees—or in children, I might add—who have always been exposed to a large number of persons. While the avoidance response is unlearned, the familiar, expected aspects of objects must be learned. The young animal must have established as residues of his experience cortical firing patterns (or cognitive structures—whichever term you like) from which new receptor inputs can be incongruous. Consider the kinds of objects regularly feared. They are, for instance, the familiar keeper or experimenter in strange clothes, the experimenter in a Hallowe'en mask, a plaster cast of a chimpanzee head (which lacks, of course, the familiarly attached body), an anesthetized chimpanzee infant (from which the familiar patterns of motion are absent). On the other hand, objects which have never entered into the young chimpanzee's life may be strange without evoking withdrawal. In other words, the feared object is one which excites receptors in a fashion which is incongruous with the central, sequential pattern of neural firing which has accrued as a residue of the chimpanzee or human infant's past experience. Until the central pattern has been learned, incongruous stimulation is impossible.

Such a conception can well account for Holmes' findings that lower-class children are less fearful than higher-class children and that boys are less fearful than girls even though both lower-class children and boys of nursery school age are likely to have had the wider experience with the sorts of situations used by Holmes to evoke fear. It may well be that being shocked and handled provides a variety of experience which leaves the rat pups which have been subjected to it less disturbed by such things as open fields and 18 hours without water, but these effects may ultimately be found to be a matter of still another mechanism. It is too early to say.

Taking seriously this incongruity-dissonance conception of the genesis of fear leads to interesting reinterpretations of a great many of the motivational phenomena of child development. Consider these few. In considering separation anxiety, the incongruity principle makes it unnecessary to puzzle about how the absence of mother could be the conditional stimulus for the traumatizing and helpless distress that has been supposed to have occurred in her absence. In considering fear of the dark, it also becomes unnecessary to puzzle about how the absence of light stimulation could so widely have been associated with painful stimulation. Multiple mothering need not be seen as a traumatizing experience in the light of this conception, but rather as an innoculation against social shyness and fear. The timidity of the overprotected child and the social shyness of the rural mountain people get an explanation which has been difficult in terms of the theory of conditioned fear.

MOTIVATION IN TERMS OF THE INCONGRUITY-DISSONANCE PRINCIPLE

This introduction of the incongruity-dissonance principle concludes the three reinterpretations I wish to present today, but I do wish to call your attention to the pervasive character of this incongruity-dissonance principle. It appears to have great explanation power which figures, in one guise or another, in several systematic theories, besides that of Hebb, all of which have been characterized as nondynamic.

Hebb's (1949) theorizing is physiological, at least in a verbal sense, in that he conceives the residues of past inputs to be stored in semi-autonomous, reverberating cerebral circuits which he terms *cell assemblies*. These cell assemblies are the neural analogue of concepts, and they get sequentially integrated into what he calls *phase sequences*. The sequential organization in time provides for the subjective phenomenon of expectation. When markedly incongruous receptor inputs disrupt this sequential organization, behavior is changed and the process is felt as unpleasant emotion. Slight degrees of incongruity, which can readily be accommodated, lend interest and may provide attractive problems, but the larger ones are repelling and perhaps even devastating.

Piaget (1952, 1954) utilizes very much the same incongruity notion to account for the development of intelligence and concepts in human children. In his system, the child comes at birth with certain sensory-motor coordinations which he terms *schemata*. Variation in stimulus situations call for adaptive *accommodations* or changes in these schemata, which changes are *assimilated* or stored as residues. Piaget also finds limited incongruities between central schemata and receptor inputs to be interesting and facilitative of growth, but incongruities which extend beyond the child's capacity for accommodation instigate withdrawal or fear and even terror. In Piaget's theory the child's gestalt-like conceptions of reality (space, time, and number) are schemata which develop through a continuous process of accommodations and assimilations and become fixed or static only when the child's schemata come to correspond so well with reality that no further accommodations are required. Here agreement among people is dictated by reality.

Helson (1947, 1948) has called the residues of immediate past experience in the typical psychophysical experiment an *adaptation level*. Both he and McClelland (1953) have seen affective arousal to be a matter of the size of the discrepancy between receptor inputs and the adaptation level. Small discrepancies may be attractively pleasant, large ones repellingly unpleasant. As an example, some of you will readily recall having experienced the affective startle that comes when you have been set to pick up what you thought was a full pail, only to find it empty.

Festinger (1957) has recently written a book entitled *A Theory of Cognitive Dissonance* in which he shows that a discrepancy between belief about a situation and perception of that situation acts like a drive. The subject acts to reduce the *dissonance* by either withdrawing from the incredible situation or by changing his beliefs, and, not incidentally, he finds the dissonance highly unpleasant.

Rogers (1951) has described the basis for anxiety as discrepancy between the "phenomenological field" and the perceived reality as represented by his two circles. Rogers' phenomenological field, however, is not the perceptually-given phenomenal field of such German phenomenologists as Delthei and Husserl. It is rather the inferred storehouse of past experience and represented in the present by expectations, aspirations, self-concept, and the like. Thus, his conceptual scheme appears to fall within the domain of the incongruity-dissonance principle.

Kelly's (1955) *Psychology of Personal Constructs* also makes central use of this principle. The term *personal constructs* refers to the ways in which individuals construe and anticipate events. These each person derives from the way in which he has experienced such events in the past. When a person's constructions fail to predict events, this is disturbing, even anxiety-producing, and it motivates some kind of change, but the change may take place in defenses against such change of constructs or in avoiding such events, or in the constructs themselves.

Perhaps, it is worth noting in closing that this incongruity-dissonance principle makes both motivation and reinforcement intrinsic to the organism's relations with its environment, intrinsic, if you will, to the organism's information-processing. It is as if the organism operated like an error-actuated, feedback system where the error is derived from discrepancy between receptor-inputs of the present and the residues of past experience which serve as the basis for anticipating the future. The dominant view of the past half century has seen both motivation and reinforcement as extrinsic to the information-processing. This has put a tremendous burden of responsibility for the management of affective motivation on parents, teachers, and all those in positions of authority and control. Visions of man completely controlled, as exemplified by George Orwell's *1984*, are conceivable only by assuming that the extrinsic motivating forces of homeostatic need and painful stimulation are completely dominant. In this light the terror of the baby chimp at seeing his keeper in a Hallowe'en mask and the irritation of the believer when his beliefs are disconfirmed are perhaps symbols of hope. They may justify Abraham Lincoln's well-known dictum that "you can fool some of the people all the time, and all the people some of the time, but you cannot fool all the people all the time."

To return to the cartoon of the lad who was improperly motivated: Perhaps, the task of developing proper motivation is best seen, at least in nutshell form, as limiting the manipulation of extrinsic factors to that minimum of keeping homeostatic need and exteroceptive drive low, in favor of facilitating basic information-processing to maximize accurate anticipation of reality.

REFERENCES

BEACH, F. A. Current concepts of play in animals. *Amer. Naturalist,* 1945, *79,* 523–541.

BEXTON, W. H., HERON, W., & SCOTT, T. H. Effects of decreased variation in the sensory environment. *Canad. J. Psychol.,* 1954, *8,* 70–76.

DAVIS, R. C., & BUCHWALD, A. M. An exploration of somatic response patterns: stimulus and sex differences. *J. comp. physiol. Psychol.,* 1957, *50,* 44–52.

DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.

FESTINGER, L. *A theory of cognitive dissonance.* Evanston, Ill.: Row, Peterson, 1957.

FREUD, S. Three contributions to the theory of sex (1905). In *The basic writings of Sigmund Freud.* (Translated by A. A. Brill) New York: Modern Library, 1938. Pp. 553–629.

FREUD, S. Instincts and their vicissitudes (1915). In *Collected papers.* Vol. IV. London: Hogarth, 1950. Pp. 60–83.

FREUD, S. *Inhibition, symptom, and anxiety* (1926). (Translated by H. A. Bunker as *The problem of anxiety.*) New York: Norton, 1936.

GRANIT, R. *Receptors and sensory perception.* New Haven: Yale University Press, 1955.

HARLOW, H. F. Learning and satiation of response in intrinsically motivated complex puzzle performance by monkeys. *J. comp. physiol. Psychol.,* 1950, *43,* 289–294.

HARLOW, H. F. The nature of love. *Amer. Psychologist,* 1958, *13,* 673–685.

HARLOW, H. F., HARLOW, M. K., & MEYER, D. R. Learning motivated by a manipulation drive. *J. exp. Psychol.,* 1950, *40,* 228–234.

HEBB, D. O. On the nature of fear. *Psychol. Rev.,* 1946, *53,* 259–276.

HEBB, D. O. *The organization of behavior.* New York: Wiley, 1949.

HEBB, D. O. Drives and the CNS (conceptual nervous system). *Psychol. Rev.,* 1955, *62,* 243–254.

HELSON, H. Adaptation-level as frames of reference for prediction of psychophysical data. *Amer. J. Psychol.,* 1947, *60,* 1–29.

HELSON, H. Adaptation-level as a basis for a quantitative theory of frame of reference. *Psychol. Rev.,* 1948, *55,* 297–313.

HOLMES, FRANCES B. An experimental study of the fears of young children. In A. T. Jersild & Frances B. Holmes, Children's fears. *Child Develpm. Monogr.*, 1935, *20*, 167–296.

HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

HUNT, J. McV. The effects of infant feeding-frustration upon adult hoarding in the albino rat. *J. abnorm. soc. Psychol.*, 1941, *36*, 338–360.

HUNT, J. McV., SCHLOSBERG, H., SOLOMON, R. L., & STELLAR, E. Studies on the effects of infantile experience on adult behavior in rats: I. Effects of infantile feeding-frustration on adult hoarding. *J. comp. physiol. Psychol.*, 1947, *40*, 291–304.

JASPER, H. H. Electrical signs of cortical activity. *Psychol. Bull.*, 1937, *34*, 411–481.

JONES, H. E., & JONES, MARY C. A study of fear. *Child Educ.*, 1928, *5*, 136–143.

KELLY, G. A. *The psychology of personal constructs.* New York: Norton, 1955.

LACEY, J. I., & LACEY, BEATRICE C. The relationship of resting autonomic activity to motor impulsivity. In *The brain and human behavior.* Baltimore: Williams & Wilkins, 1958. Pp. 144–209.

LEVINE, S., CHEVALIER, J. A., & KORCHIN, S. J. The effects of shock and handling in infancy on later avoidance learning. *J. Pers.*, 1956, *24*, 475–493.

LILLY, J. C. Mental effects of reduction of ordinary levels of physical stimuli on intact, healthy persons. *Psychiat. Res. Rep.*, 1956, No. 5, 1–9.

LINDSLEY, D. B. Psychophysiology and motivation. In M. R. Jones (Ed.), *Nebraska symposium on motivation.* Lincoln: University of Nebraska Press, 1957. Pp. 44–105.

McCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

MILLER, N. E., & DOLLARD, J. *Social learning and imitation.* New Haven: Yale University Press, 1941.

MUSSEN, P. H., & CONGER, J. J. *Child development and personality.* New York: Harper, 1956.

OLDS, J., & MILNER, P. Positive reinforcement produced by electrical stimulation of septal area and other regions of the rat brain. *J. comp. physiol. Psychol.*, 1954, *47*, 419–427.

PIAGET, J. *The origins of intelligence in children.* New York: International Universities Press, 1952.

PIAGET, J. *The construction of reality in the child.* (Translated by Margaret Cook) New York: Basic Books, 1954.

PROSSER, C. L. Action potentials in the nervous system of the crayfish: I. Spontaneous impulses. *J. cell. comp. Physiol.*, 1934, *4*, 185–209.

ROGERS, C. R. *Client-centered therapy.* Boston: Houghton Mifflin, 1951.

WATSON, J. B. *Psychological care of the infant and child.* New York: Norton, 1928.

WATSON, J. B., & RAYNOR, ROSALIE. Conditional reactions. *J. exp. Psychol.*, 1920, *3*, 1–4.

2

Motivation Reconsidered: The Concept of Competence

Robert W. White

The changing scene in motivation theory is indicated not only in the Hunt article but also in this paper by White. Both of the authors also stress the goal-seeking aspects of behavior. The selection printed here reviews the research on motivation and offers a synthesis in White's concept of "competence."

When parallel trends can be observed in realms as far apart as animal behavior and psychoanalytic ego psychology, there is reason to suppose that we are witnessing a significant evolution of ideas. In these two realms, as in psychology as a whole, there is evidence of deepening discontent with theories of motivation based upon drives. Despite great differences in the language and concepts used to express this discontent, the theme is everywhere the same: Something important is left out when we make drives the operating forces in animal and human behavior.

The chief theories against which the discontent is directed are those of Hull and of Freud. In their respective realms, drive-reduction theory and psychoanalytic instinct theory, which are basically very much alike, have acquired a considerable air of orthodoxy. Both views have an appealing simplicity, and both have been argued long enough so that their main outlines are generally known. In decided contrast is the position of those who are not satisfied with drives and

---

From *Psychological Review*, 1959, *66*: 297–323

instincts. They are numerous, and they have developed many pointed criticisms, but what they have to say has not thus far lent itself to a clear and inclusive conceptualization. Apparently there is an enduring difficulty in making these contributions fall into shape.

In this paper I shall attempt a conceptualization which gathers up some of the important things left out by drive theory. To give the concept a name I have chosen the word *competence*, which is intended in a broad biological sense rather than in its narrow everyday meaning. As used here, competence will refer to an organism's capacity to interact effectively with its environment. In organisms capable of but little learning, this capacity might be considered an innate attribute, but in the mammals and especially man, with their highly plastic nervous systems, fitness to interact with the environment is slowly attained through prolonged feats of learning. In view of the directedness and persistence of the behavior that leads to these feats of learning, I consider it necessary to treat competence as having a motivational aspect, and my central argument will be that the motivation needed to attain competence cannot be wholly derived from sources of energy currently conceptualized as drives or instincts. We need a different kind of motivational idea to account fully for the fact that man and the higher mammals develop a competence in dealing with the environment which they certainly do not have at birth and certainly do not arrive at simply through maturation. Such an idea, I believe, is essential for any biologically sound view of human nature. . . .

DEVELOPMENTS IN GENERAL PSYCHOLOGY

If a systematic survey were in order, it would be easy to show a [discontent with drive theories of motivation]. . . . Among theorists of personality, for example, something like drive orthodoxy is to be found in the work of Dollard and Miller (1950), who have translated the main concepts of Freud's psychoanalysis, including processes such as repression and displacement, into the language of reinforcement theory. With them we might put Mowrer (1950), whose searching analysis of fear as an acquired drive has led him to postulate anxiety-reduction as the master motive behind the development of the ego. Discontent with drive orthodoxy has long been expressed by Allport (1937, 1946), who not only argues for a functional autonomy of motives from their infantile roots in primary drives but also seriously questions the law of effect, the very cornerstone of reinforcement theory. Little comfort for the orthodox can be found in Murray's (1938) detailed taxonomy of needs, especially when it comes to needs such as achievement and construction, which can be tied to primary drives only by conceptual acrobatics. Murray and Kluckhohn (1953), moreover, have made a case for pleasure in activity for its own sake, reviving the *Funktionslust* proposed many years ago by Karl Buhler (1924) and recently developed in some detail by French (1952). They also argue for intrinsic mental needs: "the infant's mind is not acting most of the time as the instrument of some urgent animal drive, but is preoccupied with *gratifying itself*." Murphy (1947) takes the view that all tissues can become seats of tension and thus participants in drive; in addition to visceral drives, he postulates two independent forms, activity drives and sensory drives. Then there are workers such as Goldstein (1939) who approach the whole problem with a holistic philosophy which precludes the dictatorship of any isolated or partial drives. Goldstein (1940) assumes one master tendency, that toward self-actualization, of which the so-called visceral drives are but partial and not really isolated expressions, and which can find expression also in an urge toward perfection—toward completing what is incomplete, whether it be an outside task or the mastery of some function such as walking. It has been shown by the Ansbachers (1956) that Adler, never a friend of instinct orthodoxy, in his later years reached an idea very similar to the urge toward perfection. Maslow (1954, 1955), too, belongs with the heterodox. He insists that we should take account of growth motivation as well as the deficiency motivation implied in the visceral drives, and he offers the valuable idea of a hierarchy of motives, according to which the satisfaction of "lower" needs makes it possible for "higher" needs to emerge and become regnant in behavior.

Mention of these names must suffice here to show that the trends observed in animal psychology and psychoanalytic ego psychology are pervasive in contemporary psychological thought. Doubtless the same controversies and problems could be pointed out in child development, in cognitive psychology, and in other fields. But in order to advance to my main theme, I shall select only certain developments which bear directly on the concept of competence.

Needs for Excitement and Novelty

Human experience provides plentiful evidence of the importance of reducing excessive levels of tension. Men under wartime stress, men under pressure of pain and extreme deprivation, men with excessive work loads or too much exposure to confusing social interactions, all act as if their nervous systems craved that utterly unstimulated condition which Freud once sketched as the epitome of neural bliss. But if these same men be granted their nirvana they soon become miserable and begin to look around for a little excitement. Human experience testifies that boredom is a bad state of affairs about which something must be done. Hebb (1949) has been particularly insistent in reminding us that many of our activities, such as reading detective stories, skin-diving, or driving cars at high speeds, give clear evidence of a need to raise the level of stimulation and excitement. Men and animals alike seem at times bent on increasing the impact of the environment and even on creating mild degrees of frustration and fear. Hebb and Thompson (1954) reflect upon this as follows:

> Such phenomena are, of course, well known in man: in the liking for dangerous sports or roller coasters, where fear is deliberately courted, and in the addiction to bridge or golf or solitaire, vices whose very existence depends upon the level of difficulty of the problems presented and an optimal level of frustration. Once more, when we find such attitudes toward fear and frustration in animals, we have a better basis for supposing that we are dealing with something fundamental if a man prefers skis to the less dangerous snowshoes, or when we observe an unashamed love of work (problem solving and frustration included) in the scientist, or in the businessman who cannot retire. Such behavior in man is usually accounted for as a search for prestige, but the animal data make this untenable. It seems much more likely that solving problems and running mild risks are inherently rewarding, or, in more general terms, that the animal will always act so as to produce an optimal level of excitation (Hebb and Thompson, 1954, p. 551).

The concept of optimal stimulation has been developed by Leuba (1955), who sees it as helpful in resolving some of the problems of learning theory. Believing that most theorizing about motivation has been based upon "powerful biological or neurotic drives," Leuba bids us look at the much more common learning situations of nursery, playground, and school, where "actions which increase stimulation and produce excitement are strongly reinforced, sometimes to the dismay of parents and teachers." He proposes that there is an optimal level of stimulation, subject to variation at different times, and that learning is associated with movement toward this optimal level, downward when stimulation is too high and upward when it is too low. A similar idea is expressed by McReynolds (1956) concerning the more restricted concept of "rate of perceptualization." Monotonous conditions provide too low a rate, with boredom; excessive stimulation produces too high a rate, with disruptive excitement; the optimal rate yields the experience of pleasure. These ideas are now amply supported by recent experimental work on sensory deprivation (Lilly, 1956; Hebb, 1958).

In recent papers Young (1949, 1955) has argued for an hedonic theory of motivation, one in which affective processes "constitute a form of primary motivation." According to Young's theory, "an organism behaves so as to maximize positive affective arousal (delight, enjoyment) and to minimize negative arousal (distress)." McClelland (1953) has offered a version of hedonic theory which is of particular value in understanding the significance of novelty. Affective arousal occurs when a stimulus pattern produces a discrepancy from the existing adaptation level. Small discrepancies produce pleasant affect and a tendency to approach; large ones produce unpleasantness and a tendency toward avoidance. The child at play, like the young chimpanzee and the exploring rat, needs frequent novelty in the stimulus field in order to keep up his interest—in order to maintain pleasant discrepancies from whatever adaptation level he has reached. Hebb's (1949) theory of the neurological correlates of learning also deals with novelty, though in a somewhat different way. He equates sustained interest with a state of neural affairs in which "phase sequences" are relatively complex and are growing, in the sense of establishing new internal relations. Such a state follows most readily from a stimulus field characterized by difference-in-sameness; that is, containing much that is familiar along with certain features that are novel. If the field is entirely familiar, phase sequences run off quickly, are short-circuited, and thus fail to produce sustained interest. Hebb's theory, which has the engaging quality of being able to explain why we enjoy reading a detective story once but not right over again, expresses in a neurological hypothesis the familiar fact that well-learned, habituated processes do not in themselves greatly interest us. Interest seems to require elements of unfamiliarity: of something still to be found out and of learning still to be done.

It seems to me that these contributions, though differing as to details, speak with unanimity on their central theme and would force us, if nothing else did, to reconsider seriously the whole problem of motivation. Boredom, the unpleasantness of monotony, the attraction of novelty, the tendency to vary behavior rather than repeating it rigidly, and the seeking of stimulation and mild excitement stand as inescapable facts of human experience and clearly have their parallels in animal behavior. We may seek rest and minimal stimulation at the end of the day, but that is not what we are looking for the next morning. Even when its primary needs are satisfied and its homeostatic chores are done, an organism is alive, active, and up to something.

Dealing with the Environment

If we consider things only from the viewpoint of affect, excitement, and novelty, we are apt to overlook another important aspect of behavior, its effect upon the environment. Moving in this direction, Diamond (1939) invites us to consider the motivational properties of the sensorineural system, the apparatus whereby higher animals "maintain their relations to the environment." He conceives of this system as demanding stimulation and as acting in such a manner as to "force the environment to stimulate it." Even if one thinks only of the infant's exploring eyes and hands, it is clear that the main direction of behavior is by no means always that of reducing the impact of stimulation. When the eyes follow a moving object, or when the hand grasps an object which it has touched, the result is to preserve the stimulus and to increase its effect. In more elaborate explorations the consequence of a series of actions may be to vary the manner in which a stimulus acts upon the sense organs. It is apparent that the exploring, manipulating child produces by his actions precisely what Hebb's theory demands as a basis for continuing interest: he produces differences-in-sameness in the stimulus field.

In a critical analysis of Freud's views on the reality principle, Charlotte Bühler (1954) makes a strong case for positive interests in the environment, citing as evidence the responsiveness and adaptiveness of the newborn baby as well as the exploratory tendencies of later months. The problem is worked out in more detail by Schachtel (1954) in a paper on focal attention. Acts of focal attention are characteristically directed at particular objects, and they consist of several sustained approaches "aimed at active mental grasp" while excluding the rest of the field. These qualities can be observed even in the infant's early attempts to follow a moving object with his eyes, and they show more clearly in his later endeavors to learn how objects are related both to himself and to one another. Such behavior bespeaks "a relatively autonomous capacity for object interest." Schachtel makes the proposal that this interest is pursued precisely at those times when major needs are in abeyance. High pressure of need or anxiety is the enemy of exploratory play and is a condition, as every scientists should know, under which we are unlikely to achieve an objective grasp of the environment. Low need pressure is requisite if we are to perceive objects as they are, in their constant character, apart from hopes and fears we may at other times attach to them. Schachtel doubts that "the wish for need-satisfaction alone would ever lead to object perception and to object-oriented thought." Hence an autonomous capacity to be interested in the environment has great value for the survival of a species.

Being interested in the environment implies having some kind of satisfactory interaction with it. Several workers call attention to the possibility that satisfaction might lie in having an effect upon the environment, in dealing with it, and changing it in various ways. Gross (1901), in his classical analysis of play, attached great importance to the child's "joy in being a cause," as shown in making a clatter, "hustling things about," and playing in puddles where large and dramatic effects can be produced. "We demand a knowledge of effects," he wrote, "and to be ourselves the producers of effects." Piaget (1952) remarks upon the child's special interest in objects that are affected by his own movements. This aspect of behavior occupies a central place in the work of Skinner (1953), who describes it as "operant" and who thus "emphasizes the fact that the behavior *operates* upon the environment to generate consequences." These consequences are fed back through the sense organs and may serve to reinforce behavior even when no organic needs are involved. A rat will show an increased tendency to press a bar when this act produces a click or a buzz. A baby will continue to investigate when his efforts produce rattling or tinkling sounds or sparkling reflections from a shiny object. The young chimpanzees in Welker's experiment spent the longest time over objects which could be lighted or made to emit sounds. Skinner finds it "difficult, if not impossible, to trace these reinforcing effects to a history of conditioning." "We may plausibly argue," he continues, "that a capacity to be reinforced by any feedback from the environment would be biologically advantageous, since it would prepare the organism to manipulate the environment successfully before a given state of deprivation developed."

Woodworth's Behavior-Primacy Theory

The most far-reaching attempt to give these aspects of behavior a systematic place in the theory of motivation is contained in Woodworth's *Dynamics of Behavior* (1958). Woodworth takes his start from the idea that a great deal of human behavior appears to be directed toward producing effects upon the environment without immediate service to any aroused organic need. "Its incentives and rewards are in the field of behavior and not in the field of homeostasis." This is illustrated by exploratory behavior, which is directed outward toward the environment.

> Its long-range value as the means of making the child acquainted with the world he has to deal with later, and so equipping him through play for the serious business of life, can scarcely lie within the little child's horizon. His goals are more limited and direct: to see this or that object more closely, to find what is behind an obstacle, to hear the noise an object makes when it strikes the floor, to be told the name of a thing or person (Woodworth, 1958, p. 78).

More complex play, such as building with blocks, illustrates the same outgoing tendency and reveals more plainly the element of finding out what one can and cannot do with objects. Even social play falls into the pattern. Playmates do not chiefly supply affection or satisfy organic needs; rather, they "afford the opportunity to do something interesting in the environment."

Woodworth draws a contrast between *need-primacy* theories of motivation and the *behavior-primacy* theory. The latter holds that "all behavior is directed primarily toward dealing with the environment." It is to be noted that "dealing with the enviroment" means a good deal more than receiving stimuli and making responses. Stimuli must be taken as indicators of objects in space, and responses must be adapted to produce effects upon these objects. Even the so-called "mental" capacities, such as memory and ideational thinking, become in time high-level methods of dealing with the environment. Woodworth leaves no doubt as to what he considers basic in motivation. "We are making the claim that this direction of receptive and motor activity toward the environment is the fundamental tendency of animal and human behavior and that it is the all-pervasive primary motivation of behavior." Organic drives have to break into this constantly flowing stream of activity and turn it in a special direction. But the goals of drives cannot be achieved without effective action upon one's surroundings. The ever-present, ever-primary feature of motivation is the tendency to deal with the environment.

It may appear to some workers that Woodworth has overshot the mark by making primary what has commonly been regarded as secondary, and by reducing the familiar drives to what sounds a little like a subordinate station. Woodworth's theory, however, like Goldstein's concept of self-actualization, probably should be construed not as an attempt to downgrade the drives but rather as an insistence that they be kept in the context of a whole living organism which during its waking hours is more or less constantly active. Woodworth's emphasis on dealing with the environment makes his theory a point of culmination for many of those driftings away from drive orthodoxy which we have found to be persistent in so many different areas of psychology. It will soon appear that the concept of competence, to which I now turn, represents in many respects a similar way of thinking. It emphasizes dealing with the environment, and it belongs in the trend away from drive *orthodoxy*, but it is not intended to supplant, or even to subsume, such dynamic forces as hunger, sex, aggression, and fear, which everyone knows to be of huge importance in animal and human nature.

COMPETENCE AND THE PLAY OF CONTENTED CHILDREN

A backward glance at our survey shows considerable agreement about the kinds of behavior that are left out or handled poorly by theories of motivation based wholly on organic drives. Repeatedly we find reference to the familiar series of learned skills which starts with sucking, grasping, and visual exploration and continues with crawling and walking, acts of focal attention and perception, memory, language and thinking, anticipation, the exploring of novel places and objects, effecting stimulus changes in the environment, manipulating and exploiting the surroundings, and achieving higher levels of motor and mental coordination. These aspects of behavior have long been the province of child psychology, which has attempted to measure the slow course of their development and has shown how heavily their growth depends upon learning. Collectively they are sometimes referred to as adaptive mechanisms or as ego processes, but on the whole we are not accustomed to cast a single name over the diverse feats whereby we learn to deal with the environment.

I now propose that we gather the various kinds of behavior just mentioned, all of which have to do with effective interaction with the environment, under the general heading of competence. According to Webster, competence means fitness or ability, and the suggested synonyms include capability, capacity, efficiency, proficiency, and skill. It is therefore a suitable word to describe such things as grasping and exploring, crawling and walking, attention and perception, language and thinking, manipulating and changing the surroundings, all of which promote an effective—a competent—interaction with the environment. It is true, of course, that maturation plays a part in all these developments, but this part is heavily overshadowed by learning in all the more complex accomplishments like speech or skilled manipulation. I shall argue that it is necessary to make competence a motivational concept; there is a *competence motivation* as well as competence in its more familiar sense of achieved capacity. The behavior that leads to the building up of effective grasping, handling, and letting go of objects, to take one example, is not random behavior produced by a general overflow of energy. It is directed, selective, and persistent, and it is continued not because it serves primary drives, which indeed it cannot serve until it is almost perfected, but because it satisfies an intrinsic need to deal with the environment.

No doubt it will at first seem arbitrary to propose a single motivational conception in connection with so many and such diverse kinds of behavior. What do we gain by attributing motivational unity to such a large array of activities? We could, of course, say that each developmental sequence, such as learning to grasp or to walk, has its own built-in bit of motivation—its "aliment," as Piaget (1952) has expressed it. We could go further and say that each item of behavior has its intrinsic motive—but this makes the concept of motivation redundant. On the other hand, we might follow the lead of the animal psychologists and postulate a limited number of broader motives under such names as curiosity, manipulation, and mastery. I believe that the idea of a competence motivation is more adequate than any of these alternatives and that it points to very vital common properties which have been lost from view amidst the strongly analytical tendencies that go with detailed research.

In order to make this claim more plausible, I shall now introduce some specimens of playful exploration in early childhood. I hope that these images will serve to fix and dramatize the concept of competence in the same way that other images—the hungry animal solving problems, the child putting his finger in the candle flame, the infant at the breast, the child on the toilet, and the youthful Oedipus caught in a hopeless love triangle—have become memorable focal points for other concepts. For this purpose I turn to Piaget's (1952) studies of the growth of intelligence from its earliest manifestations in his own three children. The examples come from the first year of life, before language and verbal concepts begin to be important. They therefore represent a practical kind of intelligence which may be quite similar to what is developed by the higher animals.

As early as the fourth month, the play of the gifted Piaget children began to be "centered on a result produced in the external environment," and their behavior could be described as "rediscovering the movement which by chance exercised an advantageous action upon things" (1952, p. 151). Laurent, lying in his bassinet, learns to shake a suspended rattle by pulling a string that hangs from it. He discovers this result fortuitously before vision and prehension are fully coordinated. Let us now observe him a little later when he has reached the age of three months and ten days.

> I place the string, which is attached to the rattle, in his right hand, merely unrolling it a little so that he may grasp it better. For a moment nothing happens. But at the first shake due to chance movement of his hand, the reaction is immediate: Laurent starts when looking at the rattle and then violently strikes his right hand alone, as if he felt the resistance and the effect. The operation lasts fully a quarter of an hour, during which Laurent emits peals of laughter (Piaget, 1952, p. 162).

Three days later the following behavior is observed:

> Laurent, by chance, strikes the chain while sucking his fingers. He grasps it and slowly displaces it while looking at the rattles. He then begins to swing it very gently, which produces a slight movement of the hanging rattles and an as yet faint sound inside them. Laurent then definitely increases by degrees his own movements. He shakes the chain more and more vigorously and laughs uproariously at the result obtained (Piaget, 1952, p. 185).

Very soon it can be observed that procedures are used "to make interesting spectacles last." For instance, Laurent is shown a rubber monkey which he has not seen before. After a moment of surprise, and perhaps even fright, he calms down and makes movements of pulling the string, a procedure which has no effect in this case, but which previously has caused interesting things to happen. It is to be noticed that "interesting spectacles" consist of such things as new toys, a tin box upon which a drumming noise can be made, an unfolded newspaper, or sounds made by the observer such as snapping the fingers. Commonplace as they are to the adult mind, these spectacles enter the infant's experience as novel and apparently challenging events.

Moving ahead to the second half of the first year, we can observe behavior in which the child explores the properties of objects and tries out his repertory of actions upon them. This soon leads to active experimentation in which the child attempts to provoke new results. Again we look in upon Laurent, who has now reached the age of nine months. On different occasions he is shown a variety of new objects—for instance a notebook, a beaded purse, and a wooden parrot. His carefully observing father detects four stages of response: *(a)* visual exploration, passing the object from hand to hand, folding the purse, etc.; *(b)* tactile exploration, passing the hand all over the object, scratching, etc.; *(c)* slow moving of the object in space; *(d)* use of the repertory of action: shaking the object, striking it, swinging it, rubbing it against the side of the bassinet, sucking it, etc., "each in turn with a sort of prudence as though studying the effect produced" (1952, p. 255).

Here the child can be described as applying familiar tactics to new situations, but in a short while he will advance to clear patterns of active experimentation. At 10 months and 10 days Laurent, who is unfamiliar with bread as a nutritive substance, is given a piece for examination. He manipulates it, drops it many times, breaks off fragments and lets them fall. He has often done this kind of thing before, but previously his attention has seemed to be centered on the act of letting go. Now "he watches with great interest the body in motion; in particular, he looks at it for a long time when it has fallen, and picks it up when he can." On the following day he resumes his research.

> He grasps in succession a celluloid swan, a box, and several other small objects, in each case stretching out his arm and letting them fall. Sometimes he stretches out his arm vertically, sometimes he holds it obliquely in front of or behind his eyes. When the object falls in a new position (for example on his pillow) he lets it fall two or three times more on the same place, as though to study the spatial relation; then he modifies the situation. At a certain moment the swan falls near his mouth; now he does not suck it (even though this object habitually serves this purpose), but drops it three times more while merely making the gesture of opening his mouth (Piaget, 1952, p. 269).

These specimens will furnish us with sufficient images of the infant's use of his spare time. Laurent, of course, was provided by his studious father with a decidedly enriched environment, but no observant parent will question the fact that babies often act this way during those periods of their waking life when hunger, erotic needs, distresses, and anxiety seem to be exerting no particular pressure. If we consider this behavior under the historic headings of psychology we shall see that few processes are missing. The child gives evidence of sensing, perceiving, attending, learning, recognizing, probably recalling, and perhaps thinking in a rudimentary way. Strong emotion is lacking, but the infant's smiles, gurgles, and occasional peals of laughter strongly suggest the presence of pleasant effect. Actions appear in an organized form, particularly in the specimens of active exploration and experimentation. Apparently the child is using with a certain coherence nearly the whole repertory of psychological processes except those that accompany stress. It would be arbitrary indeed to say that one was more important than another.

These specimens have a meaningful unity when seen as transactions between the child and his environment, the child having some influence upon the environment and the environment some influence upon the child. Laurent appears to be concerned about what he can do with the chain and rattles, what he can accomplish by his own effort to reproduce and to vary the entertaining sounds. If his father observed correctly, we must add that Laurent seems to have varied his actions systematically, as if testing the effect of different degrees of effort upon the bit of environment represented by the chain and rattles. Kittens make a similar study of parameters when delicately using their paws to push pencils and other objects ever nearer to the edge of one's desk. In all such examples it is clear that the child or animal is by no means at the mercy of transient stimulus fields. He selects for continuous treatment those aspects of his environment which he finds it possible to affect in some way. His behavior is selective, directed, persistent—in short, motivated.

Motivated toward what goal? In these terms, too, the behavior exhibits a little of everything. Laurent can be seen as appeasing a stimulus hunger, providing his sensorium with an agreeable level of stimulation by eliciting from the environment a series of interesting sounds, feels, and sights. On the other hand we might emphasize a need for activity and see him as trying to reach a pleasurable level of neuromuscular exercise. We can also see another possible goal in the behavior: the child is achieving knowledge, attaining a more differentiated cognitive map of his environment and thus satisfying an exploratory tendency or motive of curiosity. But it is equally possible to discern a theme of mastery, power, or control, perhaps even a bit of primitive self-assertion, in the child's concentration upon those aspects of the environment which respond in some way to his own activity. It looks as if we had found too many goals, and perhaps our first impulse is to search for some key to tell us which one is really important. But this, I think, is a mistake that would be fatal to understanding.

We cannot assign priority to any of these goals without pausing arbitrarily in the cycle of transaction between child and environment and saying, "This is the real point." I propose instead that the real point is the transactions as a whole. If the behavior gives satisfaction, this satisfaction is not associated with a particular moment in the cycle. It does not lie solely in sensory stimulation, in a bettering of the cognitive map, in coordinated action, in motor exercise, in a feeling of effort and of effects produced, or in the appreciation of change brought about in the sensory field. These are all simply aspects of a process which at this stage has to be conceived as a whole. The child appears to be occupied with the agreeable task of developing an effective familiarity with his environment. This involves discovering the effects he can have on the environment and the effects the environment will have on him. To the extent that these results are preserved by learning, they build up an increased competence in dealing with the environment. The child's play can thus be viewed as serious business, though to him it is merely something that is interesting and fun to do.

Bearing in mind these examples, as well as the dealings with environment pointed out by other workers, we must now attempt to describe more fully the possible nature of the motivational aspect of competence. It needs its own name, and in view of the foregoing analysis I propose that this name be *effectance*.

EFFECTANCE

The new freedom produced by two decades of research on animal drives is of great help in this undertaking. We are no longer obliged to look for a source of energy external to the nervous system, for a consummatory climax, or for a fixed connection between reinforcement and tension-reduction. Effectance motivation cannot, of course, be conceived as having a source in tissues external to the nervous system. It is in no sense a deficit motive. We must assume it to be neurogenic, its "energies" being simply those of the living cells that make up the nervous system. External stimuli play an important part, but in terms of "energy" this part is secondary, as one can see most clearly when environmental stimulation is actively sought. Putting it picturesquely, we might say that the effectance urge represents what the neuromuscular system wants to do when it is otherwise unoccupied or is gently stimulated by the environment. Obviously there are no consummatory acts; satisfaction would appear to lie in the arousal and maintaining of activity rather than in its slow decline toward bored passivity. The motive need not be conceived as intense and powerful in the sense that hunger, pain, or fear can be powerful when aroused to high pitch. There are plenty of instances in which children refuse to leave their absorbed play in order to eat or to visit the toilet. Strongly aroused drives, pain, and anxiety, however, can be conceived as overriding the effectance urge and capturing the energies of the neuromuscular system. But effectance motivation is persistent in the sense that it regularly occupies the spare waking time between episodes of homeostatic crisis.

In speculating upon this subject we must bear in mind the continuous nature of behavior. This is easier said than done; habitually we break things down in order to understand them, and such units as the reflex arc, the stimulus-response sequence, and the single transaction with the environment seem like inevitable steps toward clarity. Yet when we apply such an analysis to playful exploration we lose the most essential aspect of the behavior. It is constantly circling from stimulus to perception to action to effect to stimulus to perception, and so on around; or, more properly, these processes are all in continuous action and continuous change. Dealing with the environment means carrying on a continuing transaction which gradually changes one's relation to the environment. Because there is no consummatory climax, satisfaction has to be seen as lying in a considerable series of transactions, in a trend of behavior rather than a goal that is achieved. It is difficult to make the word "satisfaction" have this connotation, and we shall do well to replace it by "feeling of efficacy" when attempting to indicate the subjective and affective side of effectance.

It is useful to recall the findings about novelty: the singular effectiveness of novelty in engaging interest and for a time supporting persistent behavior. We also need to consider the selective continuance of transactions in which the animal or child has a more or less pronounced effect upon the environment—in which something happens as a consequence of his activity. Interest is not aroused and sustained when the stimulus field is so familiar that it gives rise at most to reflex acts or automatized habits. It is not sustained when actions produce no effects or changes in the stimulus field. Our conception must therefore be that effectance motivation is aroused by stimulus conditions which offer, as Hebb (1949) puts it, difference-in-sameness. This leads to variability and novelty of response, and interest is best sustained when the resulting action affects the stimulus so as to produce further difference-in-sameness. Interest wanes when action begins to have less effect; effectance motivation subsides when a situation has been explored to the point that it no longer presents new possibilities.

We have to conceive further that the arousal of playful and exploratory interest means the appearance of organization involving both the cognitive and active aspects of behavior. Change in the stimulus field is not an end in itself, so to speak; it happens when one is passively moved about, and it may happen as a consequence of random movements without becoming focalized and instigating exploration. Similarly, action which has effects is not an end in itself, for if one unintentionally kicks away a branch while walking, or knocks something off a table, these effects by no means necessarily become involved in playful investigation. Schachtel's (1954) emphasis on focal attention becomes helpful at this point. The playful and exploratory behavior shown by Laurent is not random or casual. It involves focal *attention* to some object—the fixing of some aspect of the stimulus field so that it stays relatively constant—and it also involves the focalizing of *action* upon this object. As Diamond (1939) has expressed it, response under these conditions is "relevant to the stimulus," and it is change in the *focalized* stimulus that so strongly affects the level of interest. Dealing with the environment means directing focal attention to some part of it and organizing actions to have some effect on this part.

In our present state of relative ignorance about the workings of the nervous system it is impossible to form a satisfactory idea of the neural basis of effectance motivation, but it should at least be clear that the concept does not refer to any and every kind of neural action. It refers to a particular kind of activity, as inferred from particular kinds of behavior. We can say that it does not include reflexes and other kinds of automatic response. It does not include well-learned, automatized patterns, even those that are complex and highly organized. It does not include behavior in the service of effectively aroused drives. It does not even include activity that is highly random and discontinuous, though such behavior may be its most direct forerunner. The urge toward competence is inferred specifically from behavior that shows a lasting focalization and that has the characteristics of exploration and experimentation, a kind of variation within the focus. When this particular sort of activity is aroused in the nervous system, effectance motivation is being aroused, for it is characteristic of this particular sort of activity that it is selective, directed, and persistent, and that instrumental acts will be learned for the sole reward of engaging in it.

Some objection may be felt to my introducing the word *competence* in connection with behavior that is so often playful. Certainly the playing child is doing things for fun, not because of a desire to improve his competence in dealing with the stern hard world. In order to forestall misunderstanding, it should be pointed out that the usage here is parallel to what we do when we connect sex with its biological goal of reproduction. The sex drive aims for pleasure and gratification, and reproduction is a consequence that is presumably unforeseen by animals and by man at primitive levels of understanding. Effectance motivation similarly aims for the feeling of efficacy, not for the vitally important learnings that come as its consequence. If we consider the part played by competence motivation in adult human life we can observe the same parallel. Sex may now be completely and purposefully divorced from reproduction but nevertheless pursued for the pleasure it can yield. Similarly, effectance motivation may lead to continuing exploratory interests or active adventures when in fact there is no longer any gain in actual competence or any need for it in terms of survival. In both cases the motive is capable of yielding surplus satisfaction well beyond what is necessary to get the biological work done.

In infants and young children it seems to me sensible to conceive of effectance motivation as undifferentiated. Later in life it becomes profitable to distinguish various motives such as cognizance, construction, mastery, and achievement. It is my view that all such motives have a root in effectance motivation. They are differentiated from it through life experiences which emphasize one or another aspect of the cycle of transaction with environment. Of course, the motives of later childhood and of adult life are no longer simple and can almost never be referred to a single root. They can acquire loadings of anxiety, defense, and compensation, they can become fused with unconscious fantasies of a sexual, aggressive, or omnipotent character, and they can gain force because of their service in producing realistic results in the way of income and career. It is not my intention to cast effectance in the star part in adult motivation. The acquisition of motives is a complicated affair in which simple and sovereign theories grow daily more obsolete. Yet it may be that the satisfaction of effectance contributes significantly to those feelings of interest which often sustain us so well in day-to-day actions, particularly when the things we are doing have continuing elements of novelty.

THE BIOLOGICAL SIGNIFICANCE OF COMPETENCE

The conviction was expressed at the beginning of this paper that some such concept as competence, interpreted motivationally, was essential for any biologically sound view of human nature. This necessity emerges when we consider the nature of living systems, particularly when we take a longitudinal view. What an organism does at a given moment does not always give the right clue as to what it does over a period of time. Discussing this problem, Angyal (1941) has proposed that we should look for the general pattern followed by the total organismic process over the course of time. Obviously this makes it necessary to take account of growth. Angyal defines life as "a process of self-expansion"; the living system "expands at the expense of its surroundings," assimilating parts of the environment and transforming them into functioning parts of itself. Organisms differ from other things in nature in that they are "self-governing entities" which are to some extent "autonomous." Internal processes govern them as well as external "heteronomous" forces. In the course of life there is a relative increase in the preponderance of internal over external forces. The living system expands, assimilates more of the environment, transforms its surroundings so as to bring them under greater control. "We may say," Angyal writes, "that the general dynamic trend of the organism is toward an increase of autonomy. . . . The human being has a characteristic tendency toward self-determination, that is, a tendency to resist external influences and to subordinate the heteronomous forces of the physical and social environment to its own sphere of influence." The trend toward increased autonomy is characteristic so long as growth of any kind is going on, though in the end the living system is bound to succumb to the pressure of heteronomous forces.

Of all living creatures, it is man who takes the longest strides toward autonomy. This is not because of any unusual tendency toward bodily expansion at the expense of the environment. It is rather that man, with his mobile hands and abundantly developed brain, attains an extremely high level of competence in his transactions with his surroundings. The building of houses, roads and bridges, the making of tools and instruments, the domestication of plants and animals—all qualify as planful changes made in the environment so that it comes more or less under control and serves our purposes rather than intruding upon them. We meet the fluctuations of outdoor temperature, for example, not only with our bodily homeostatic mechanisms, which alone would be painfully unequal to the task, but also with clothing, buildings, controlled fires, and such complicated devices as self-regulating central heating and air conditioning. Man as a species has developed a tremendous power of bringing the environment into his service, and each individual member of the species must attain what is really quite an impressive level of competence if he is to take part in the life around him.

We are so accustomed to these human accomplishments that it is hard to realize how long an apprenticeship they require. At the outset the human infant is a slow learner in comparison with other animal forms. Hebb (1949) speaks of "the astonishing inefficiency of man's first learning, as far as immediate results are concerned," an inefficiency which he attributes to the large size of the association areas in the brain and the long time needed to bring them under sensory control. The human lack of precocity in learning shows itself even in comparison with one of the next of kin: as Hebb points out, "the human baby takes six months, the chimpanzee four months, before making a clear distinction between friend and enemy." Later in life the slow start will pay dividends. Once the fundamental perceptual elements, simple associations, and conceptual sequences have been established, later learning can proceed with ever-increasing swiftness and complexity. In Hebb's words, "learning at maturity concerns patterns and events whose parts at least are familiar and which already have a number of other associations."

This general principle of cumulative learning, starting from slowly acquired rudiments and proceeding thence with increasing efficiency, can be illustrated by such processes as manipulation and locomotion, which may culminate in the acrobat devising new stunts or the dancer working out a new ballet. It is especially vivid in the case of language, where the early mastery of words and pronunciation seems such a far cry from spontaneous adult speech. A strong argument has been made by Hebb (1949) that the learning of visual forms proceeds over a similar course from slowly learned elements to rapidly combined patterns. Circles and squares, for example, cannot be discriminated at a glance without a slow apprenticeship involving eye movements, successive fixations, and recognition of angles. Hebb proposes that the recognition of visual patterns without eye movement "is possible only as the result of an intensive and prolonged visual training that goes on from the moment of birth, during every moment that the eyes are open, with an increase in skill evident over a period of 12 to 16 years at least."

On the motor side there is likewise a lot to be cumulatively learned. The playing, investigating child slowly finds out the relationships between what he does and what he experiences. He finds out, for instance, how hard he must push what in order to produce what effect. Here the S-R formula is particularly misleading. It would come nearer the truth to say that the child is busy learning R-S connections—the effects that are likely to follow upon his own behavior. But even in this reversed form the notion of bonds or connections would still misrepresent the situation, for it is only a rare specimen of behavior that can properly be conceived as determined by fixed neural channels and a fixed motor response. As Hebb has pointed out, discussing the phenomenon of "motor equivalence" named by Lashley (1942), a rat which has been trained to press a lever will press it with the left forepaw, the right forepaw, by climbing upon it, or by biting it; a monkey will open the lid of a food box with either hand, with a foot, or even with a stick; and we might add that a good baseball player can catch a fly ball while running in almost any direction and while in almost any posture, including leaping in the air and plunging forward to the ground. All of these feats are possible because of a history of learnings in which the main lesson has been the effects of actions upon the stimulus fields that represent the environment. What has been learned is not a fixed connection but a flexible relationship between stimulus fields and the effects that can be produced in them by various kinds of action.

One additional example, drawn this time from Piaget (1952), is particularly worth mentioning because of its importance in theories of development. Piaget points out that a great deal of mental development depends upon the idea that the world is made up of objects having substance and permanence. Without such an "object concept" it would be impossible to build up the ideas of space and causality and to arrive at the fundamental distinction between self and external world. Observation shows that the object concept, "far from being innate or ready-made in experience, is constructed little by little." Up to the age of seven and eight months the Piaget children searched for vanished objects only in the sense of trying to continue the actions, such as sucking or grasping, in which the objects had played a part. When an object was really out of sight or touch, even if only because it was covered by a cloth, the infants undertook no further exploration. Only gradually, after some study of the displacement of objects by moving, swinging, and dropping them, does the child begin to make an active search for a vanished object, and only still more gradually does he learn, at 12 months or more, to make allowance for the object's sequential displacements and thus to seek it where it has gone rather than where it was last in sight. Thus it is only through cumulative learning that the child arrives at the idea of permanent substantial objects.

The infant's play is indeed serious business. If he did not while away his time pulling strings, shaking rattles, examining wooden parrots, dropping pieces of bread and celluloid swans, when would he learn to discriminate visual patterns, to catch and throw, and to build up his concept of the object? When would he acquire the many other foundation stones necessary for cumulative learning? The more closely we analyze the behavior of the human infant, the more clearly do we realize that infancy is not simply a time when the nervous system matures and the muscles grow stronger. It is a time of active and continuous learning, during which the basis is laid for all those processes, cognitive and motor, whereby the child becomes able to establish effective transactions with his environment and move toward a greater degree of autonomy. Helpless as he may seem until he begins to toddle, he has by that time already made substantial gains in the achievement of competence.

Under primitive conditions survival must depend quite heavily upon achieved competence. We should expect to find things so arranged as to favor and maximize this achievement. Particularly in the case of man, where so little is provided innately and so much has to be learned through experience, we should expect to find highly advantageous arrangements for securing a steady cumulative learning about the properties of the environment and the extent of possible transactions. Under these circumstances we might expect to find a very powerful drive operating to insure progress toward competence, just as the vital goals of nutrition and reproduction are secured by powerful drives, and it might therefore seem paradoxical that the interests of competence should be so much entrusted to times of play and leisurely exploration. There is good reason to suppose, however, that a strong drive would be precisely the wrong arrangement to secure a flexible knowledgeable power of transaction with the environment. Strong drives cause us to learn certain lessons well, but they do not create maximum familiarity with our surroundings.

This point was demonstrated more than half a century ago in some experiments by Yerkes and Dodson (1908). They showed that maximum motivation did not lead to the most rapid solving of problems, especially if the problems were complex. For each problem there was an optimum level of motivation, neither the highest nor the lowest, and the optimum was lower for more complex tasks. The same problem has been discussed more recently by Tolman (1948) in his paper on cognitive maps. A cognitive map can be narrow or broad, depending upon the range of cues picked up in the course of learning. Tolman suggests that one of the conditions which tend to narrow the range of cues is a high level of motivation. In everyday terms, a man hurrying to an important business conference is likely to perceive only the cues that help him to get there faster, whereas a man taking a stroll after lunch is likely to pick up a substantial amount of casual information about his environment. The latent learning experiments with animals, and experiments such as those of Johnson (1953) in which drive level has been systematically varied in a situation permitting incidental learning, give strong support to this general idea. In a recent contribution, Bruner, Matter, and Papanek (1955) make a strong case for the concept of breadth of learning and provide additional evidence that it is favored by moderate and hampered by strong motivation. The latter "has the effect of speeding up learning at the cost of narrowing it." Attention is concentrated upon the task at hand and little that is extraneous to this task is learned for future use.

These facts enable us to see the biological appropriateness of an arrangement which uses periods of less intense motivation for the development of competence. This is not to say that the narrower but efficient learnings that go with the reduction of strong drives make no contribution to general effectiveness. They are certainly an important element in capacity to deal with the environment, but a much greater effectiveness results from having this capacity fed also from learnings that take place in quieter times. It is then that the infant can attend to matters of lesser urgency, exploring the properties of things he does not fear and does not need to eat, learning to gauge the force of his string-pulling when the only penalty for failure is silence on the part of attached rattles, and generally accumulating for himself a broad knowledge and a broad skill in dealing with his surroundings.

The concept of competence can be most easily discussed by choosing, as we have done, examples of interaction with the inanimate environment. It applies equally well, however, to transactions with animals and with other human beings, where the child has the same problem of finding

Acc. No. 2320
Date 23-2-72

136.5
GOR

R.T., West Ben
Library

out what effects he can have upon the environment and what effects it can have upon him. The earliest interactions with members of the family may involve needs so strong that they obscure the part played by effectance motivation, but perhaps the example of the well-fed baby diligently exploring the several features of his mother's face will serve as a reminder that here, too, there are less urgent moments when learning for its own sake can be given free rein.

In this closing section I have brought together several ideas which bear on the evolutionary significance of competence and of its motivation. I have sought in this way to deepen the biological roots of the concept and thus help it to attain the stature in the theory of behavior which has not been reached by similar concepts in the past. To me it seems that the most important proving ground for this concept is the effect it may have on our understanding of the development of personality. Does it assist our grasp of early object relations, the reality principle, and the first steps in the development of the ego? Can it be of service in distinguishing the kinds of defense available at different ages and in providing clues to the replacement of primitive defenses by successful adaptive maneuvers? Can it help fill the yawning gap known as the latency period, a time when the mastery of school subjects and other accomplishments claim so large a share of time and energy? Does it bear upon the self and the vicissitudes of self-esteem, and can it enlighten the origins of psychological disorder? Can it make adult motives and interests more intelligible and enable us to rescue the concept of sublimation from the difficulties which even its best friends have recognized? I believe it can be shown that existing explanations of development are not satisfactory and that the addition of the concept of competence cuts certain knots in personality theory. But this is not the subject of the present communication, where the concept is offered much more on the strength of its logical and biological probability.

REFERENCES

ALLPORT, G. W. *Personality: A psychological interpretation.* New York: Holt, 1937.

ALLPORT, G. W. Effect: A secondary principle of learning. *Psychol. Rev.,* 1946, *53,* 335–347.

ANGYAL, A. *Foundations for a science of personality.* New York: Commonwealth Fund, 1941.

ANSBACHER, H. L., & ANSBACHER, R. R. (EDS.) *The individual psychology of Alfred Adler.* New York: Basic Books, 1956.

BRUNER, J. S., MATTER, J., & PAPANEK, M. L. Breadth of learning as a function of drive level and mechanization. *Psychol. Rev.,* 1955, *62,* 1–10.

BÜHLER, C. The reality principle. *Amer. J. Psychother.,* 1954, *8,* 626–647.

BÜHLER, K. *Die geistige Entwicklung des Kendes.* 4th ed. Jena: Gustav Fischer, 1924.

DIAMOND, S. A neglected aspect of motivation. *Sociometry,* 1939, *2,* 77–85.

DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.

FRENCH, T. M. *The integration of behavior.* Vol. I. *Basic postulates.* Chicago: University of Chicago Press, 1952.

GOLDSTEIN, K. *The organism.* New York: American Book, 1939.

GOLDSTEIN, K. *Human nature in the light of psychopathology.* Cambridge, Mass.: Harvard University Press, 1940.

GROSS, K. *The play of man.* (Translated by E. L. Baldwin) New York: D. Appleton, 1901.

HEBB, D. O. *The organization of behavior.* New York: Wiley, 1949.

HEBB, D. O. The motivating effects of exteroceptive stimulation. *Amer. Psychologist,* 1958, *13,* 109–113.

HEBB, D. O., & THOMPSON, W. R. The social significance of animal studies. In G. Lindzey (Ed.), *Handbook of social psychology.* Vol. I. Cambridge, Mass.: Addison-Wesley, 1954. Pp. 532–561.

JOHNSON, E. E. The role of motivational strength in latent learning. *J. comp. physiol. Psychol.,* 1953, *45,* 526–530.

LASHLEY, K. S. The problem of cerebral organization in vision. In H. Klüver, *Visual mechanisms.* Lancaster, Pa.: Jaques Cattell, 1942. Pp. 301–322.

LEUBA, C. Toward some integration of learning theories: The concept of optimal stimulation. *Psychol. Rep.,* 1955, *1,* 27–33.

LILLY, J. C. Mental effects of reduction of ordinary levels of physical stimuli on intact, healthy persons. *Psychiat. res. Rep.,* 1956, No. 5, 1–9.

MASLOW, A. H. *Motivation and personality.* New York: Harper, 1954.

MASLOW, A. H. Deficiency motivation and growth motivation. In M. R. Jones (Ed.), *Nebraska symposium on motivation.* Lincoln: University of Nebraska Press, 1955. Pp. 1–30.

McCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century, 1953.

McREYNOLDS, P. A restricted conceptualization of human anxiety and motivation. *Psychol. Rep.*, 1956, *2*, 293–312. Monogr. Suppl. 6.

MOWRER, O. H. *Learning theory and personality dynamics.* New York: Ronald, 1950.

MURPHY, G. *Personality: A biosocial approach to origins and structure.* New York: Harper, 1947.

MURRAY, H. A. *Explorations in personality.* New York & London: Oxford University Press, 1938.

MURRAY, H. A., & KLUCKHOHN, C. Outline of a conception of personality. In C. Kluckhohn, H. A. Murray, & D. M. Schneider (Eds.), *Personality in nature, society, and culture.* 2nd. ed. New York: Knopf, 1953.

PIAGET, J. *The origins of intelligence in children.* (Translated by Margaret Cook) New York: International Universities Press, 1952.

SCHACHTEL, E. G. The development of focal attention and the emergence of reality. *Psychiatry*, 1954, *17*, 309–324.

SKINNER, B. F. *Science and human behavior.* New York: Macmillan, 1953.

TOLMAN, E. C. Cognitive maps in rats and men. *Psychol. Rev.*, 1948, *55*, 189–208.

WELKER, W. L. Some determinants of play and exploration in chimpanzees. *J. comp. physiol. Psychol.*, 1956, *49*, 84–89.

WOODWORTH, R. S. *Dynamics of behavior.* New York: Holt, 1958.

YERKES, R. M., & DODSON, J. D. The relation of strength of stimulus to rapidity of habit-formation. *J. comp. neurol. Psychol.*, 1908, *18*, 459–482.

YOUNG, P. T. Food-seeking drive, affective process, and learning. *Psychol. Rev.*, 1949, *56*, 98–121.

YOUNG, P. T. The role of hedonic processes in motivation. In M. R. Jones (Ed.), *Nebraska symposium on motivation.* Lincoln: University of Nebraska Press, 1955. Pp. 193–238.

3

Three Faces of Intellect

J. P. Guilford

The changing concepts of the nature of man and the new techniques for measuring his capabilities are illustrated in Guilford's work. His factor-analytic approach to personality and his current use of this approach to intelligence have broadened our understanding of both areas. Much of the recent work in creativity uses the conceptualization of convergent and divergent thinking depicted in his model.

My subject is in the area of human intelligence, in connection with which the names of Terman and Stanford have become known the world over. The Stanford Revision of the Binet intelligence scale has been the standard against which all other instruments for the measurement of intelligence have been compared. The term IQ or intelligence quotient has become a household word in this country. This is illustrated by two brief stories.

A few years ago, one of my neighbors came home from a PTA meeting, remarking: "That Mrs. So-And-So, thinks she knows so much. She kept talking about the 'intelligence *quota*'; imagine. Why, everybody knows that IQ stands for 'intelligence *quiz*.'"

The other story comes from a little comic strip in a Los Angeles morning newspaper, called "Junior Grade." In the first picture a little boy meets a little girl, both apparently about the first-grade level. The little girl remarks, "I have a high IQ." The little boy, puzzled, said, "You have a what?" The little girl repeated, "I have a high IQ," then went on her way. The little boy, looking thoughtful, said, "And she looks like such a nice little girl, too."

It is my purpose to speak about the analysis of this thing called human intelligence into its components. I do not believe that either Binet or Terman, if they were still with us, would object to the idea of a searching and detailed study of intelligence, aimed toward a better understanding of its nature. Preceding the development of his intelligence scale, Binet had done much research on different kinds of thinking activities and apparently recognized that intelligence has a number of aspects. It is to the lasting credit of both Binet and Terman that they introduced such a great variety of tasks into their intelligence scales.

Two related events of very recent history make it imperative that we learn all we can regarding the nature of intelligence. I am referring to the advent of the artificial satellites and planets and to the crisis in education that has arisen in part as a consequence. The preservation of our way of life and our future security depend upon our most important national resources: our intellec-

---

From *American Psychologist*, 1959, *14*: 469–479.

tual abilities and, more particularly, our creative abilities. It is time, then, that we learn all we can about those resources.

Our knowledge of the components of human intelligence has come about mostly within the last 25 years. The major sources of this information in this country have been L. L. Thurstone and his associates, the wartime research of psychologists in the United States Air Forces, and more recently the Aptitudes Project at the University of Southern California, now in its tenth year of research on cognitive and thinking abilities. The results from the Aptitudes Project that have gained perhaps the most attention have pertained to creative-thinking abilities. These are mostly novel findings. But to me, the most significant outcome has been the development of a unified theory of human intellect, which organizes the known, unique or primary intellectual abilities into a single system called the "structure of intellect." It is to this system that I shall devote the major part of my remarks, with very brief mentions of some of the implications for the psychology of thinking and problem solving, for vocational testing, and for education.

The discovery of the components of intelligence has been by means of the experimental application of the method of factor analysis. It is not necessary for you to know anything about the theory or method of factor analysis in order to follow the discussion of the components. I should like to say, however, that factor analysis has no connection with or resemblance to psychoanalysis. A positive statement would be more helpful, so I will say that each intellectual component or factor is a unique ability that is needed to do well in a certain class of tasks or tests. As a general principle we find that certain individuals do well in the tests of a certain class, but they may do poorly in the tests of another class. We conclude that a factor has certain properties from the features that the tests of a class have in common. I shall give you very soon a number of examples of tests, each representing a factor.

THE STRUCTURE OF INTELLECT

Although each factor is sufficiently distinct to be detected by factor analysis, in very recent years it has become apparent that the factors themselves can be classified because they resemble one another in certain ways. One basis of classification is according to the basic kind of process or operation performed. This kind of classification gives us five major groups of intellectual abilities: factors of cognition, memory, convergent thinking, divergent thinking, and evaluation.

Cognition means discovery or rediscovery or recognition. Memory means retention of what is cognized. Two kinds of productive-thinking operations generate new information from known information and remembered information. In divergent-thinking operations we think in different directions, sometimes searching, sometimes seeking variety. In convergent thinking the information leads to one right answer or to a recognized best or conventional answer. In evaluation we reach decisions as to goodness, correctness, suitability, or adequacy of what we know, what we remember, and what we produce in productive thinking.

A second way of classifying the intellectual factors is according to the kind of material or content involved. The factors known thus far involve three kinds of material or content: the content may be figural, symbolic, or semantic. Figural content is concrete material such as is perceived through the senses. It does not represent anything except itself. Visual material has properties such as size, form, color, location, or texture. Things we hear or feel provide other examples of figural material. Symbolic content is composed of letters, digits, and other conventional signs, usually organized in general systems, such as the alphabet or the number system. Semantic content is in the form of verbal meanings or ideas, for which no examples are necessary.

When a certain operation is applied to a certain kind of content, as many as six general kinds of products may be involved. There is enough evidence available to suggest that, regardless of the combinations of operations and content, the same six kinds of products may be found associated. The six kinds of products are: units, classes, relations, systems, transformations, and implications. So far as we have determined from factor analysis, these are the only fundamental kinds of products that we can know. As such, they may serve as basic classes into which one might fit all kinds of information psychologically.

The three kinds of classifications of the factors of intellect can be represented by means of a single solid model, shown in Figure 1. In this model, which we call the "structure of intellect," each dimension represents one of the modes of variation of the factors.[1] Along one dimension are found the various kinds of operations, along a second one are the various kinds of products, and along the third are various kinds of content. Along the dimension of content a fourth category has been added, its kind of content being designated as "behavioral." This category has been added on a purely theoretical basis to represent the general area sometimes called "social intelligence." More will be said about this section of the model later.

---

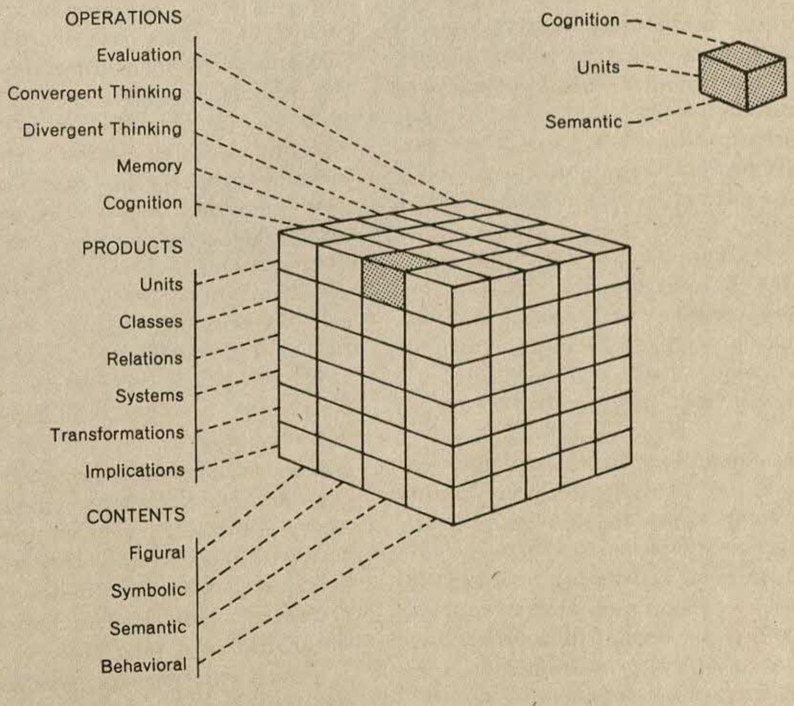
[1] For an earlier presentation of the concept, see Guilford (1956)

In order to provide a better basis for understanding the model and a better basis for accepting it as a picture of human intellect, I shall do some exploring of it with you systematically, giving some examples of tests. Each cell in the model calls for a certain kind of ability that can be described in terms of operation, content, and product, for each cell is at the intersection of a unique combination of kinds of operation, content, and product. A test for that ability would have the same three properties. In our exploration of the model, we shall take one vertical layer at a time, beginning with the front face. The first layer provides us with a matrix of 18 cells (if we ignore the behavioral column for which there are as yet no known factors) each of which should contain a cognitive ability.

The Cognitive Abilities

We know at present the unique abilities that fit logically into 15 of the 18 cells for cognitive abilities. Each row presents a triad of similar abilities, having a single kind of product in common. The factors of the first row are concerned with the knowing of units. A good test of the ability to cognize figural units is the Street Gestalt Completion Test. In this test, the recognition of familiar pictured objects in silhouette form is made difficult for testing purposes by blocking out parts of those objects. There is another factor that is known to involve the perception of auditory figures—in the form of melodies, rhythms, and speech sounds—and still another factor involving kinesthetic forms. The presence of three factors in one cell (they are conceivably distinct abilities, although this has not been tested) suggests that more generally, in the figural column, at least, we should expect

Fig. 1: *A cubical model representing the structure of intellect.*



to find more than one ability. A fourth dimension pertaining to variations in sense modality may thus apply in connection with figural content. The model could be extended in this manner if the facts call for such an extension.

The ability to cognize symbolic units is measured by tests like the following:

Put vowels in the following blanks to make real words:

P____W____R
M____RV____L
C____RT____N

Rearrange the letters to make real words:

RACIH
TVOES
KLCCO

The first of these two tests is called Disemvoweled Words, and the second Scrambled Words.

The ability to cognize semantic units is the well-known factor of verbal comprehension, which is best measured by means of a vocabulary test, with items such as:

GRAVITY means ________
CIRCUS means ________
VIRTUE means ________

From the comparison of these two factors it is obvious that recognizing familiar words as letter structures and knowing what words mean depend upon quite different abilities.

For testing the abilities to know classes of units, we may present the following kinds of items, one with symbolic content and one with semantic content:

Which letter group does not belong?
XECM PVAA QXIN VTRO

Which object does not belong?
clam tree oven rose

A figural test is constructed in a completely parallel form, presenting in each item four figures, three of which have a property in common and the fourth lacking that property.

The three abilities to see relationships are also readily measured by a common kind of test, differing only in terms of content. The well-known analogies test is applicable, two items in symbolic and semantic form being:

JIRE : KIRE : : FORA :
KORE KORA LIRE GORA GIRE
poetry : prose : : dance :
music walk sing talk jump

Such tests usually involve more than the ability to cognize relations, but we are not concerned with this problem at this point.

The three factors for cognizing systems do not at present appear in tests so closely resembling one another as in the case of the examples just given. There is nevertheless an underlying common core of logical similarity. Ordinary space tests, such as Thurstone's Flags, Figures, and Cards or Part V (Spatial Orientation) of the Guilford-Zimmerman Aptitude Survey (GZAS), serve in the figural column. The system involved is an order or arrangement of objects in space. A system that uses symbolic elements is illustrated by the Letter Triangle Test, a sample item of which is:

__
d __
b e __
a c f ?

What letter belongs at the place of the question mark?

The ability to understand a semantic system has been known for some time as the factor called general reasoning. One of its most faithful indicators is a test composed of arithmetic-reasoning items. That the phase of understanding only is important for measuring this ability is shown by the fact that such a test works even if the examinee is not asked to give a complete solution; he need only show that he structures the problem properly. For example, an item from the test Necessary Arithmetical Operations simply asks what operations are needed to solve the problem:

A city lot 48 feet wide and 149 feet deep costs $79,432. What is the cost per square foot?

A. add and multiply
B. multiply and divide
C. subtract and divide
D. add and subtract
E. divide and add

Placing the factor of general reasoning in this cell of the structure of intellect gives us some new conceptions of its nature. It should be a broad ability to grasp all kinds of systems that are conceived in terms of verbal concepts, not restricted to the understanding of problems of an arithmetical type.

Transformations are changes of various kinds, including modifications in arrangement, organization, or meaning. In the figural column for the transformations row, we find the factor known as visualization. Common measuring instruments for this factor are the surface-development tests, and an example of a different kind is Part VI (Spatial Visualization) of the GZAS. A test of the ability to make transformations of meaning, for the factor in the semantic column, is called Similarities. The examinee is asked to state several ways in which two objects, such as an apple and an orange, are alike. Only by shifting the meanings of both is the examinee able to give many responses to such an item.

In the set of abilities having to do with the cognition of implications, we find that the individual goes beyond the information given, but not to the extent of what might be called drawing conclusions. We may say that he extrapolates. From the given information he expects or foresees certain consequences, for example. The two factors found in this row of the cognition matrix were first called "foresight" factors. Foresight in connection with figural material can be tested by means of paper-and-pencil mazes. Foresight in connection with ideas, those pertaining to events, for example, is indicated by a test such as Pertinent Questions:

In planning to open a new hamburger stand in a certain community, what four questions should be considered in deciding upon its location?

The more questions the examinee asks in response to a list of such problems, the more he evidently foresees contingencies.

The Memory Abilities

The area of memory abilities has been explored less than some of the other areas of operation, and only seven of the potential cells of the memory matrix have known factors in them. These cells are restricted to three rows: for units, relations, and systems. The first cell in the memory matrix is now occupied by two factors, parallel to two in the corresponding cognition matrix: visual memory and auditory memory. Memory for series of letters or numbers, as in memory span tests, conforms to the conception of memory for symbolic units. Memory for the ideas in a paragraph conforms to the conception of memory for semantic units.

The formation of associations between units, such as visual forms, syllables, and meaningful words, as in the method of paired associates, would seem to represent three abilities to remember relationships involving three kinds of content. We know of two such abilities, for the symbolic and semantic columns. The memory for known systems is represented by two abilities very recently discovered (Christal, 1958). Remembering the arrangement of objects in space is the nature of an ability in the figural column, and remembering a sequence of events is the nature of a corresponding ability in the semantic column. The differentiation between these two abilities implies that a person may be able to say where he saw an object on a page, but he might not be able to say on which of several pages he saw it after leafing through several pages that included the right one. Considering the blank rows in the memory matrix, we should expect to find abilities also to remember classes, transformations, and implications, as well as units, relations, and systems.

The Divergent-Thinking Abilities

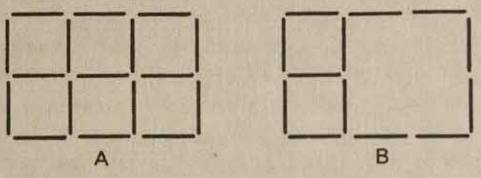
The unique feature of divergent production is that a *variety* of responses is produced. The product is not completely determined by the given information. This is not to say that divergent thinking does not come into play in the total process of reaching a unique conclusion, for it comes into play wherever there is trial-and-error thinking.

The well-known ability of word fluency is tested by asking the examinee to list words satisfying a specified letter requirement, such as words beginning with the letter "s" or words ending in "-tion." This ability is now regarded as a facility in divergent production of symbolic units. The parallel semantic ability has been known as ideational fluency. A typical test item calls for listing objects that are round and edible. Winston Churchill must have possessed this ability to a high degree. Clement Attlee is reported to have said about him recently that, no matter what problem came up, Churchill always seemed to have about ten ideas. The trouble was, Attlee continued, he did not know which was the good one. The last comment implies some weakness in one or more of the evaluative abilities.

The divergent production of class ideas is believed to be the unique feature of a factor called "spontaneous flexibility." A typical test instructs the examinee to list all the uses he can think of for a common brick, and he is given eight minutes. If his responses are: build a house, build a barn, build a garage, build a school, build a

Fig. 2: *A sample item from the test Match Problems. The problem in this item is to take away four matches and leave three squares. The solution is given.*

Item from the test Match Problems



Take away four matches in A, leaving three squares and nothing more. *Answer:* B

church, build a chimney, build a walk, and build a barbecue, he would earn a fairly high score for ideational fluency but a very low score for spontaneous flexibility, because all these uses fall into the same class. If another person said: make a door stop, make a paper weight, throw it at a dog, make a bookcase, drown a cat, drive a nail, make a red powder, and use for baseball bases, he would also receive a high score for flexibility. He has gone frequently from one class to another.

A current study of unknown but predicted divergent-production abilities includes testing whether there are also figural and symbolic abilities to produce multiple classes. An experimental figural test presents a number of figures that can be classified in groups of three in various ways, each figure being usable in more than one class. An experimental symbolic test presents a few numbers that are also to be classified in multiple ways.

A unique ability involving relations is called "associational fluency." It calls for the production of a variety of things related in a specified way to a given thing. For example, the examinee is asked to list words meaning about the same as "good" or to list words meaning about the opposite of "hard." In these instances the response produced is to complete a relationship, and semantic content is involved. Some of our present experimental tests call for the production of varieties of relations, as such, and involve figural and symbolic content also. For example, given four small digits, in how many ways can they be related in order to produce a sum of eight?

One factor pertaining to the production of systems is known as expressional fluency. The rapid formation of phrases or sentences is the essence of certain tests of this factor. For example, given the initial letters:

W——c——e——n——

with different sentences to be produced, the examinee might write "We can eat nuts" or "Whence came Eve Newton?" In interpreting the factor, we regard the sentence as a symbolic system. By analogy, a figural system would be some kind of organization of lines and other elements, and a semantic system would be in the form of a verbally stated problem or perhaps something as complex as a theory.

In the row of the divergent-production matrix devoted to transformations, we find some very interesting factors. The one called "adaptive flexibility" is now recognized as belonging in the figural column. A faithful test of it has been Match Problems. This is based upon the common game that uses squares, the sides of which are formed by match sticks. The examinee is told to take away a given number of matches to leave a stated number of squares with nothing left over. Nothing is said about the sizes of the squares to be left. If the examinee imposes upon himself the restriction that the squares that he leaves must be of the same size, he will fail in his attempts to do items like that in Figure 2. Other odd kinds of solutions are introduced in other items, such as overlapping squares and squares within squares, and so on. In another variation of Match Problems the examinee is told to produce two or more solutions for each problem.

A factor that has been called "originality" is now recognized as adaptive flexibility with semantic material, where there must be a shifting of meanings. The examinee must produce the shifts or changes in meaning and so come up with novel, unusual, clever, or farfetched ideas. The Plot Titles Test presents a short story, the examinee being told to list as many appropriate titles as he can to head the story. One story is about a missionary who has been captured by cannibals in Africa. He is in the pot and about to be boiled when a princess of the tribe obtains a promise for his release if he will become her mate. He refuses and is boiled to death.

In scoring the test, we separate the responses into two categories, clever and nonclever. Examples of nonclever responses are: African Death, Defeat of a Princess, Eaten by Savages, The Princess, The African Missionary, In Darkest Africa, and Boiled by Savages. These titles are appropriate but commonplace. The number of such responses serves as a score for ideational fluency. Examples of clever responses are: Pot's Plot, Potluck Dinner, Stewed Parson, Goil or Boil, A Mate Worse Than Death, He Left a Dish for a Pot, Chaste in Haste, and A Hot Price for Freedom. The number of clever responses given by an examinee is his score for originality, or the divergent production of semantic transformations.

Another test of originality presents a very novel task so that any acceptable response is unusual for the individual. In the Symbol Production Test the examinee is to produce a simple symbol to stand for a noun or a verb in each short sentence, in other words to invent something like pictographic symbols. Still another test of originality asks for writing the "punch lines" for cartoons, a task that almost automatically challenges the examinee to be clever. Thus, quite a variety of tests offer approaches to the measurement of originality, including one or two others that I have not mentioned.

Abilities to produce a variety of implications are assessed by tests calling for elaboration of given information. A figural test of this type provides the examinee with a line or two, to which he is to add other lines to produce an object. The more lines he adds, the greater his score. A semantic test gives the examinee the outlines of a plan to which he is to respond by stating all the details he can think of to make the plan work. A new test we are trying out in the symbolic area presents two simple equations such as $B - C = D$ and $z = A + D$. The examinee is to make as many other equations as he can from this information.

The Convergent-Production Abilities

Of the 18 convergent-production abilities expected in the three content columns, 12 are now recognized. In the first row, pertaining to units, we have an ability to name figural properties (forms or colors) and an ability to name abstractions (classes, relations, and so on). It may be that the ability in common to the speed of naming forms and the speed of naming colors is not appropriately placed in the convergent-thinking matrix. One might expect that the thing to be produced in a test of the convergent production of figural units would be in the form of figures rather than words. A better test of such an ability might somehow specify the need for one particular object, the examinee to furnish the object.

A test for the convergent production of classes (Word Grouping) presents a list of 12 words that are to be classified in four, and only four, meaningful groups, no word to appear in more than one group. A parallel test (Figure Concepts Test) presents 20 pictured real objects that are to be grouped in meaningful classes of two or more each.

Convergent production having to do with relationships is represented by three known factors, all involving the "eduction of correlates," as Spearman called it. The given information includes one unit and a stated relation, the examinee to supply the other unit. Analogies tests that call for completion rather than a choice between alternative answers emphasize this kind of ability. With symbolic content such an item might read:

pots stop bard drab rats __?__

A semantic item that measures eduction of correlates is:

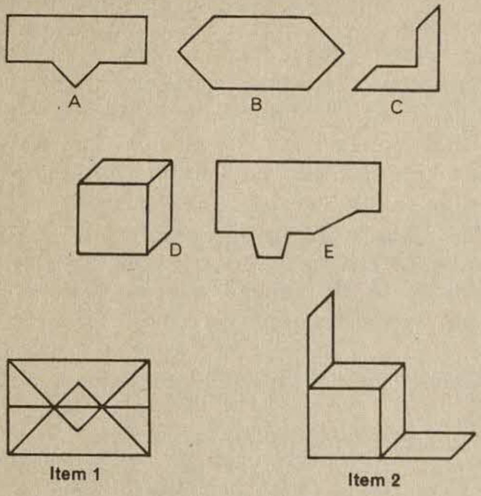
The absence of sound is ________.

Incidentally, the latter item is from a vocabulary-completion test, and its relation to the factor of ability to produce correlates indicates how, by change of form, a vocabulary test may indicate an ability other than that for which vocabulary tests are usually intended, namely, the factor of verbal comprehension.

Only one factor for convergent production of systems is known, and it is in the semantic column. It is measured by a class of tests that may be called ordering tests. The examinee may be presented with a number of events that ordinarily have a best or most logical order, the events being presented in scrambled order. The presentation may be pictorial, as in the Picture Arrangement Test, or verbal. The pictures may be taken from a cartoon strip. The verbally presented events may be in the form of the various steps needed to plant a new lawn. There are undoubtedly other kinds of systems than temporal order that could be utilized for testing abilities in this row of the convergent-production matrix.

In the way of producing transformations of a unique variety, we have three recognized factors, known as redefinition abilities. In each case, redefinition involves the changing of functions or uses of parts of one unit and giving them new functions or uses in some new unit. For testing the ability of figural redefinition, a task based upon the Gottschaldt figures is suitable. Figure 3 shows the kind of item for such a test. In recog-

Fig. 3: *Sample items from a test of Hidden Figures, based upon the Gottschaldt figures. Which of the simpler figures is concealed within each of the two more complex figures?*



nizing the simpler figure within the structure of a more complex figure, certain lines must take on new roles.

In terms of symbolic material, the following sample items will illustrate how-groups of letters in given words must be readapted to use in other words. In the test Camouflaged Words, each sentence contains the name of a sport or game:

> I did not know that he was ailing.
> To beat the Hun, tin goes a long way.

For the factor of semantic redefinition, the Gestalt Transformation Test may be used. A sample item reads:

From which object could you most likely make a needle?

A. a cabbage
B. a splice
C. a steak
D. a paper box
E. a fish

The convergent production of implications means the drawing of fully determined conclusions from given information. The well-known factor of numerical facility belongs in the symbolic column. For the parallel ability in the figural column, we have a test known as Form Reasoning, in which rigorously defined operations with figures are used. For the parallel ability in the semantic column, the factor sometimes called "deduction" probably qualifies. Items of the following type are sometimes used.

> Charles is younger than Robert
> Charles is older than Frank
> Who is older: Robert or Frank?

Evaluative Abilities

The evaluative area has had the least investigation of all the operational categories. In fact, only one systematic analytical study has been devoted to this area. Only eight evaluative abilities are recognized as fitting into the evaluation matrix. But at least five rows have one or more factors each, and also three of the usual columns or content categories. In each case, evaluation involves reaching decisions as to the accuracy, goodness, suitability, or workability of information. In each row, for the particular kind of product of that row, some kind of criterion or standard of judgment is involved.

In the first row, for the evaluation of units, the important decision to be made pertains to the identity of a unit. Is this unit identical with that one? In the figural column we find the factor long known as "perceptual speed." Tests of this factor invariably call for decisions of identity, for example, Part IV (Perceptual Speed) of the GZAS or Thurstone's Identical Forms. I think it has been generally wrongly thought that the ability involved is that of cognition of visual forms. But we have seen that another factor is a more suitable candidate for this definition and for being in the very first cell of the cognitive matrix. It is parallel to this evaluative ability but does not require the judgment of identity as one of its properties.

In the symbolic column is an ability to judge identity of symbolic units, in the form of series of letters or numbers or of names of individuals.

Are members of the following pairs identical or not:

825170493____825176493
dkeltvmpa____dkeltvmpa
C. S. Meyerson____C. E. Meyerson

Such items are common in tests of clerical aptitude.

There should be a parallel ability to decide whether two ideas are identical or different. Is the idea expressed in that one? Do these two proverbs express essentially the same idea? Such tests exist and will be used to test the hypothesis that such an ability can be demonstrated.

No evaluative abilities pertaining to classes have as yet been recognized. The abilities having to do with evaluation where relations are concerned must meet the criterion of logical consistency. Syllogistic-type tests involving letter symbols indicate a different ability than the same type of test involving verbal statements. In the figural column we might expect that tests incorporating geometric reasoning or proof would indicate a parallel ability to sense the soundness of conclusions regarding figural relationships.

The evaluation of systems seems to be concerned with the internal consistency of those systems, so far as we can tell from the knowledge of one such factor. The factor has been called "experiential evaluation," and its representative test presents items like that in Figure 4 asking "What is wrong with this picture?" The things wrong are often internal inconsistencies.

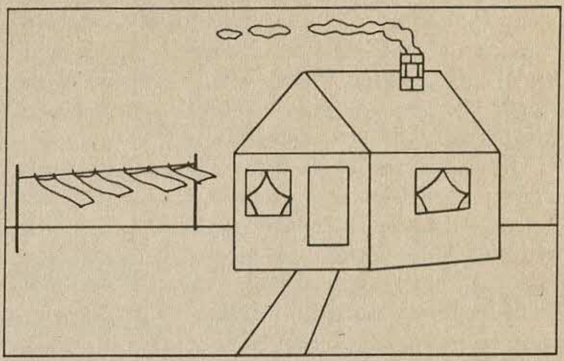
A semantic ability for evaluating transformations is thought to be that known for some time as "judgment." In typical judgment tests, the examinee is asked to tell which of five solutions to a practical problem is most adequate or wise. The solutions frequently involve improvisations, in other words, adaptations of familiar objects to unusual uses. In this way the items present redefinitions to be evaluated.

A factor known first as "sensitivity to problems" has become recognized as an evaluative ability having to do with implications. One test of the factor, the Apparatus Test, asks for two needed improvements with respect to each of several common devices, such as the telephone or the toaster. The Social Institutions Test, a measure of the same factor, asks what things are wrong with each of several institutions, such as tipping or national elections. We may say that defects or deficiencies are implications of an evaluative kind. Another interpretation would be that seeing defects and deficiencies are evaluations of implications to the effect that the various aspects of something are all right.[2]

SOME IMPLICATIONS OF THE STRUCTURE OF INTELLECT

For Psychological Theory

Fig. 4: *A sample item from the test Unusual Details. What two things are wrong with this figure?*



Although factor analysis as generally employed is best designed to investigate ways in which individuals differ from one another, in other words, to discover traits, the results also tell us much about how individuals are alike. Consequently, information regarding the factors and their interrelationships gives us understanding of functioning individuals. The five kinds of intellectual abilities in terms of operations may be said to represent five ways of functioning. The kinds of intellectual abilities distinguished according to varieties of test content and the kinds of abilities distinguished according to varieties of products suggest a classification of basic forms of information or knowledge. The kind of organism suggested by this way of looking at intellect is that of an agency for dealing with information of various kinds in various ways. The concepts provided by the distinctions among the intellectual abilities and by their classifications may be very useful in our future investigations of learning, memory, problem solving, invention, and decision making, by whatever method we choose to approach those problems.

For Vocational Testing

With about 50 intellectual factors already known, we may say that there are at least 50 ways of being intelligent. It has been facetiously suggested that there seem to be a great many more ways of being stupid, unfortunately. The structure of intellect is a theoretical model that predicts as many as 120 distinct abilities, if every cell of the model contains a factor. Already we know that two cells contain two or more factors each, and there probably are actually other cells of this type. Since the model was first conceived, 12 factors predicted by it have found places in it. There is consequently hope of filling many of the other vacancies, and we may eventually end up with more than 120 abilities.

---

[2] For further details concerning the intellectual factors, illustrative tests, and the place of the factors in the structure of intellect, see Guilford (1959)

The major implication for the assessment of intelligence is that to know an individual's intellectual resources thoroughly we shall need a surprisingly large number of scores. It is expected that many of the factors are intercorrelated, so there is some possibility that by appropriate sampling we shall be able to cover the important abilities with a more limited number of tests. At any rate, a multiple-score approach to the assessment of intelligence is definitely indicated in connection with future vocational operations.

Considering the kinds of abilities classified as to content, we may speak roughly of four kinds of intelligence. The abilities involving the use of figural information may be regarded as "concrete" intelligence. The people who depend most upon these abilities deal with concrete things and their properties. Among these people are mechanics, operators of machines, engineers (in some aspects of their work), artists, and musicians.

In the abilities pertaining to symbolic and semantic content, we have two kinds of "abstract" intelligence. Symbolic abilities should be important in learning to recognize words, to spell, and to operate with numbers. Language and mathematics should depend very much upon them, except that in mathematics some aspects, such as geometry, have strong figural involvement. Semantic intelligence is important for understanding things in terms of verbal concepts and hence is important in all courses where the learning of facts and ideas is essential.

In the hypothesized behavioral column of the structure of intellect, which may be roughly described as "social" intelligence, we have some of the most interesting possibilities. Understanding the behavior of others and of ourselves is largely nonverbal in character. The theory suggests as many as 30 abilities in this area, some having to do with understanding, some with productive thinking about behavior, and some with the evaluation of behavior. The theory also suggests that information regarding behavior is also in the form of the six kinds of products that apply elsewhere in the structure of intellect, including units, relations, systems, and so on. The abilities in the area of social intelligence, whatever they prove to be, will possess considerable importance in connection with all those individuals who deal most with other people: teachers, law officials, social workers, therapists, politicians, statesmen, and leaders of other kinds.

For Education

The implications for education are numerous, and I have time just to mention a very few. The most fundamental implication is that we might well undergo transformations with respect to our conception of the learner and of the process of learning. Under the prevailing conception, the learner is a kind of stimulus-response device, much on the order of a vending machine. You put in a coin, and something comes out. The machine learns what reaction to put out when a certain coin is put in. If, instead, we think of the learner as an agent for dealing with information, where information is defined very broadly, we have something more analogous to an electronic computor. We feed a computor information; it stores that information; it uses that information for generating new information, either by way of divergent or convergent thinking; and it evaluates its own results. Advantages that a human learner has over a computor include the step of seeking and discovering new information from sources outside itself and the step of programing itself. Perhaps even these steps will be added to computors, if this has not already been done in some cases.

At any rate, this conception of the learner leads us to the idea that learning is discovery of information, not merely the formation of associations, particularly associations in the form of stimulus-response connections. I am aware of the fact that my proposal is rank heresy. But if we are to make significant progress in our understanding of human learning and particularly our understanding of the so-called higher mental processes of thinking, problem solving, and creative thinking, some drastic modifications are due in our theory.

The idea that education is a matter of training the mind or of training the intellect has been rather unpopular, wherever the prevailing psychological doctrines have been followed. In theory, at least, the emphasis has been upon the learning of rather specific habits or skills. If we take our cue from factor theory, however, we recognize that most learning probably has both specific and general aspects or components. The general aspects may be along the lines of the factors of intellect. This is not to say that the individual's status in each factor is entirely determined by learning. We do not know to what extent each factor is determined by heredity and to what extent by learning. The best position for educators to take is that possibly every intellectual factor can be developed in individuals at least to some extent by learning.

If education has the general objective of developing the intellects of students, it can be suggested that each intellectual factor provides a particular goal at which to aim. Defined by a certain combination of content, operation, and product, each goal ability then calls for certain kinds of practice in order to achieve improvement in it. This implies choice of curriculum and the choice or invention of teaching methods that will most likely accomplish the desired results.

Considering the very great variety of abilities revealed by the factorial exploration of intellect, we are in a better position to ask whether any general intellectual skills are now being neglected in education and whether appropriate balances are being observed. It is often observed these days that we have fallen down in the way of producing resourceful, creative graduates. How true this is, in comparison with other times, I do not know. Perhaps the deficit is noticed because the demands for inventiveness are so much greater at this time. At any rate, realization that the more conspicuously creative abilities appear to be concentrated in the divergent-thinking category, and also to some extent in the transformation category, we now ask whether we have been giving these skills appropriate exercise. It is probable that we need a better balance of training in the divergent-thinking area as compared with training in convergent thinking and in critical thinking or evaluation.

The structure of intellect as I have presented it to you may or may not stand the test of time. Even if the general form persists, there are likely to be some modifications. Possibly some different kind of model will be invented. Be that as it may, the fact of a multiplicity of intellectual abilities seems well established.

There are many individuals who long for the good old days of simplicity, when we got along with one unanalyzed intelligence. Simplicity certainly has its appeal. But human nature is exceedingly complex, and we may as well face that fact. The rapidly moving events of the world in which we live have forced upon us the need for knowing human intelligence thoroughly. Humanity's peaceful pursuit of happiness depends upon our control of nature and of our own behavior; and this, in turn, depends upon understanding ourselves, including our intellectual resources.

REFERENCES

CHRISTAL, R. E. Factor analytic study of visual memory. *Psychol. Monogr.*, 1958, *72*, No. 13 (Whole No. 466).

GUILFORD, J. P. The structure of intellect. *Psychol. Bull.*, 1956, *53*, 267–293.

GUILFORD, J. P. *Personality*. New York: McGraw-Hill, 1959.

4

Individual Patterns of Development

Nancy Bayley

The contributions of the California Growth Study to our understanding of human development are monumental. This article by Bayley, who was a member of the research team for over two decades, is but a small sample of the thoroughness of the work. The articles in Part V based on other phases of the Study's work are focused upon a particular period (adolescence), but Bayley's material covers the entire first decade of life.

One of the perennial concerns of those who are trying to understand the young human organism is with the evaluation of his present status in the light of his earlier development, and as it bears on his future. We are concerned with the cause-effect relationships between various inherent factors and environmental situations, or between early manifestations and later development. We want to be able to predict, to know *how*, and how early we can tell what a child will turn out to be like.

To do this, we set up norms of development, based on the central tendencies and variabilities of healthy children. In applying to these norms the repeated measures of children over time, we have become increasingly aware that the growth

---

From *Child Development*, 1956, *27*, 45–74. Reprinted with permission of the Society for Research in Child Development.

of individuals is often unstable. In a given character, such as height, or intelligence, a child may, over a period of years, shift from high to average, to low and back to average again, as compared with his age peers. The very frequency of these shifts leads us to assume that, for the most part, they are normal and healthy patterns of growth. But we should like to understand the reasons for them, in order to insure optimal development. If we can find some condition that is operating to enhance or lower a child's efficiency or to stimulate or hinder growth, then there is the possibility of doing something to control it. But the processes of growth are very complex: it is often difficult to find a plausible event that could make the difference in the developmental trend. Furthermore, such an event may seem relevant for one child, but may make little difference in other children.

In our longitudinal studies, where we have made repeated observations on the same children over time, and charted the development of each child in relation to the norms, we find that individual patterns are the rule. It is a rare child who follows the same course in all of the observed variables through all of his growth. To explain these individual patterns, we are repeatedly being forced to re-examine the nature of development, and to reconsider the relevant conditions that can determine the course of growth in an individual organism.

If we begin by considering the early stages of the embryo and fetus, it becomes evident that the organism does not start with the discrete structures and functions that we are able to observe and classify later on in childhood. Starting with the relatively simple fertilized ovum, there is, with development, a progressive differentiation of both structures and functions, with a corresponding trend in each toward independence from the others. However, this independence is never complete in the living organism, and there are gradations of degree of independency that are correlated with the amount of similarity or difference between the two factors being compared. Thus the first reflex occurs in the embryo only after a given structural differentiation has been attained. In the neonate the observable behaviors are almost exclusively reflex in nature, and for much of the first year mental development is difficult to distinguish, first from reflex responses, and then from motor coordinations. This very condition forces on us the awareness that there is continuous change in both structural and functional organization in early development. It seems reasonable to assume that as long as there is growth or developmental change in an organism there will continue to be changes in organization; that is, the relative contribution of different factors to any "whole" structure or function will shift as each factor develops, matures and declines.

An adequate frame of reference for understanding change, or differentiation and growth, must include individual differences in temporal processes, as well as structural and functional processes. That is, not only do structure and function develop and become differentiated from each other, but also they do this at different rates. These differences in timing occur for different aspects of a single organism, as well as between different organisms.

Consequently, in studying individual patterns of growth we must take all of these dimensions of change into account. They all are important both in the predictions of a child's future development and in the evaluations of his current status. Only when we understand the regular processes of changing organization will we be able adequately to assess the relative importance of specific environmental variables in altering the processes of development in an individual.

THE STUDY OF PATTERNS OF DEVELOPMENT

When we have a series of measures and observations on a child over time, as he grows from infancy to adulthood, there are several ways in which we can orient the data, in order to study his patterns of development.

Normative Age Curves of Growth

The usual way to present growth, or change, over time is to compare a child, at successive ages, with values that have been obtained for other (presumably normal) children the same ages. This may be done by presenting individual curves of growth which show the child's scores in relation to some measures of central tendency and variation. Sometimes the normative frames of reference are curves representing the means and intervals of one standard deviation from the mean. Sometimes series of percentile curves are used such as the 10th, 25th, 50th, 75th and 90th percentiles.

In either of these methods of establishing norms the data are grouped on the basis of chronological age. That is, a child is compared with other children the *same age*. It is sometimes possible, however, to set up norms on groups that are homogeneous in other respects. For example, one could group together measurements of children who are at the same stages of maturity in that factor which is being measured. Thus children of the same skeletal age may be grouped together, as being equally mature physically. Such standards are useful in determining whether a child is developing normally for his own time schedule in that and related characteristics, regardless of his calendar age.

When we have longitudinal data on the same individuals over long periods of time, there is another method of expressing growth that has been found very informative. The child's own status at some stable or well-defined period can be made the standard against which his status at other times can be compared. For example, his height at closure of epiphyses in the bones of the hand, or his intelligence score at 21 years, can be taken as the point of reference. If we call this point 100 per cent, then the fraction of that figure can be computed for the relevant measures taken at all other ages. This procedure has the advantage of eliminating individual differences in size or in capacities. It thus makes more clear-cut the study of time patterns, or growth rates, within the individual. These time patterns can then be compared for a normal sample, to learn the normal variations in precocity or retardation, or lability of the factor under consideration.

Relative Scores and Increment Scores

The methods described so far are applied to total scores or measurements in individual functions or structures. Relative scores or increment scores can be applied to any of these general methods. Customarily, relative scores are expressed by comparison with the average at a given chronological age, as standard scores or quotients (e.g., IQ or DQ). They could, however, be related to skeletal age or some other standard of maturational status. Age curves of such indices are often charted to show stability or change in relative status over time. They do not indicate growth directly: thus the constant IQ when charted by age is represented as a straight horizontal line, regardless of increasing ability. A continuously dropping standard score may, for example, be consistent with steady slow growth in actual ability.

Where growth is more rapid at some ages than at others, the pattern of change is often charted as a series of increments, or rates of change. For example, annual increments in height have characteristic changes in rates of growth, with periods of rapid growth in infancy and again at adolescence.

All of these methods yield information that is useful in evaluating a child and the course of his growth, but each is limited, and each supplements the other. Some methods are more practical, and more readily put to use. Others, more difficult, may serve their purpose if they only indicate the allowances that must be made in the interpretation of data presented in the more customary way.

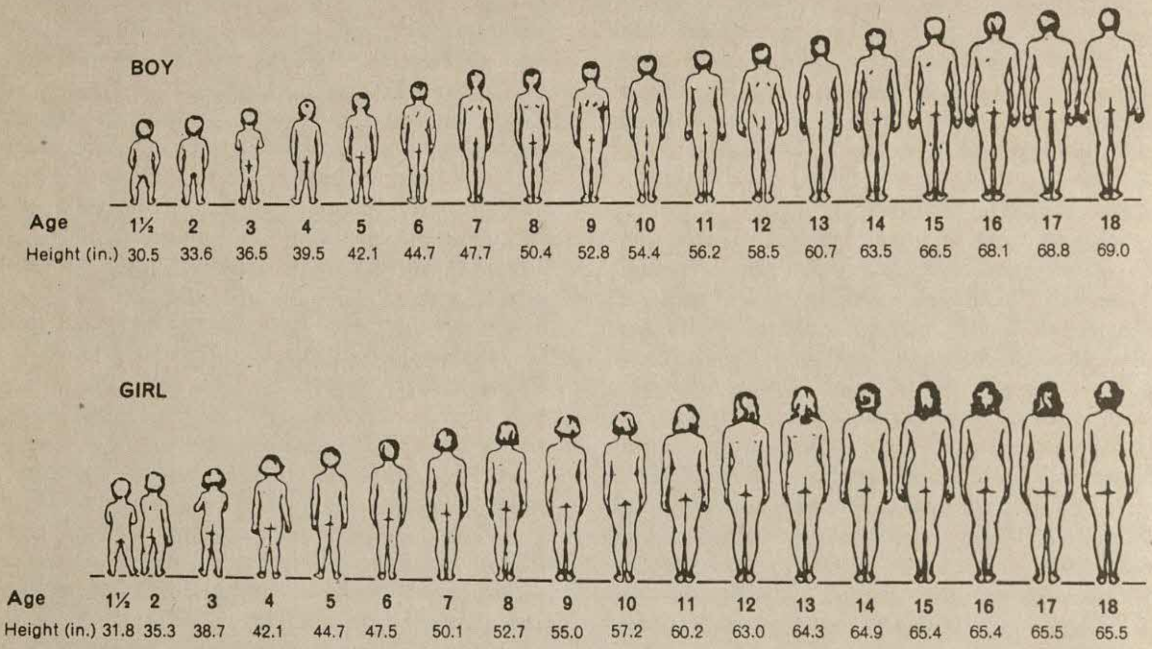
The Problem of Comparability of Units of Measurement

When we wish to study the over-all development of a child, the charting of growth curves presents another problem, that of setting up comparable units of measurement for the purpose of comparing different kinds of data. We would like to compare centimeters with kilograms, and, what is far more difficult, compare absolute measurements such as centimeters and kilograms with imprecise units such as increments in intellectual ability or changing patterns of skeletal ossification. Usually these latter have been derived from observations on normal samples, reduced to the amount of change to be expected from one age to another, and expressed as mental age or skeletal age and the like. Such age norms do not take into account the facts that growth is not usually equal in intensity or extent at all ages, nor are the standard deviations of the different types of test equal or alike in their growth trends. Thus, conversion of scores into "ages" tends to minimize some evident differences in rates of development, and to exaggerate others. In the field of intelligence testing this has been considered a very real problem, and various attempts have been made to set up "absolute" scales of mental units. Such absolute scales are usually derived by comparisons of overlapping distributions of scores at successive ages, taking into account both increments with age, and standard deviations. These scores are not really absolute, but they are useful approaches in this direction.

Factors and Proportions

The proportionate contributions of various substructures or part-functions to the whole may

Fig. 1: *Growth study of one boy and one girl from 15 months to 18 years. From the Berkeley Growth Study.*



be expressed by ratios, or other statistical devices. For example, in anthropometrics we can describe varieties of body build by relating several measures to stature: the ratio, stem length/height, differentiates between the long-bodied and the long-legged; bi-iliac diameter/height is used to separate the slender from the broad-built; weight/height also differentiates between relative stockiness and linearity of build. Statistical methods, such as covariance and factor analysis, are applied to anthropometric values and to behavior data, to show the relative contributions of several sub-scores to a total measure of physique, intelligence, or personality, and the like. All of these devices can be applied to the study of age changes in individuals by securing repeated values at successive ages wherever the relevant measures are available on the same subjects, as in longitudinal studies.

DATA FROM THE BERKELEY GROWTH STUDY

I should like now to illustrate some of these methods as applied to the Berkeley Growth Study (Jones & Bayley, 1941) data, showing their values and limitations. In the process, I hope to show, in patterns of individual development, the nature of interrelationships between several different aspects of mental and physical development from birth to young adulthood, as they are found in this small normal sample. The subjects of the Berkeley Growth Study were originally selected as normal infants when first seen in the hospital during their first few days of life. Forty of these children have continued in the study and have been tested through at least 18 and some through 25 years.

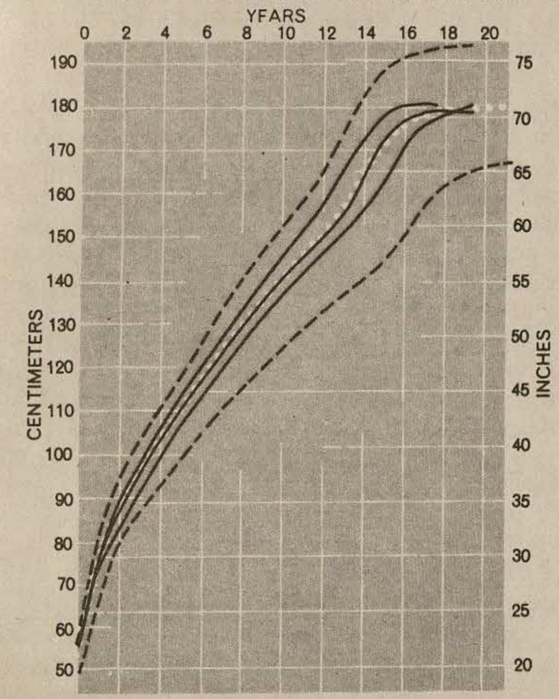
Patterns of Physical Growth .

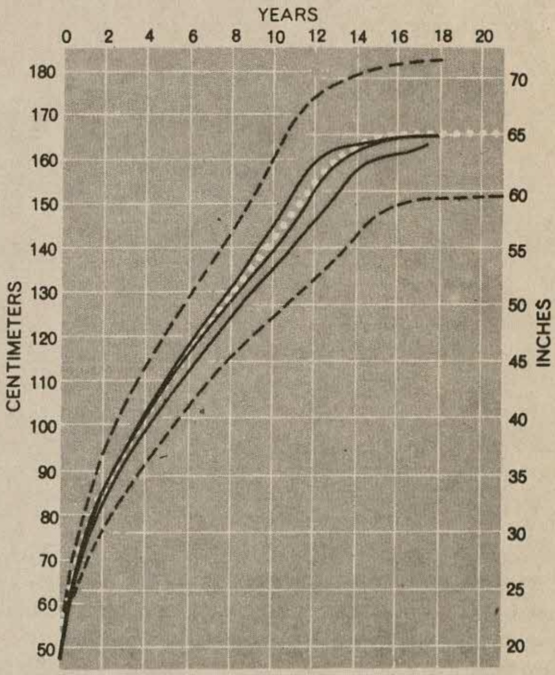
Let us start with some aspects of the actual appearance of the normal child as he grows up. In Figure 1 we have a form of growth curve of one boy and one girl from 15 months to 18 years. These are tracings from photographs, drawn to scale and spaced to that growth in size is represented at annual intervals. We see that the girl gets her height sooner, but the boy is eventually taller. Over most of this period the legs are growing faster than the stem. With maturing of the body at adolescence the girl's hips and the boy's shoulders broaden. . . . The head makes

Fig. 2: *Growth curves of boys' heights, showing the differences between the conventional method of mean height for chronological age (dotted line) and height for age in groups that are equated for physical maturity.*

Fig. 3: *Growth curves for girls' heights, showing the differences between the conventional method of mean height for chronological age (dotted line) and height for age in groups that are equated for physical maturity.*

GROWTH CURVES OF HEIGHT BY AGE


Boys Maturing at Average, Accelerated, and Retarded Rates

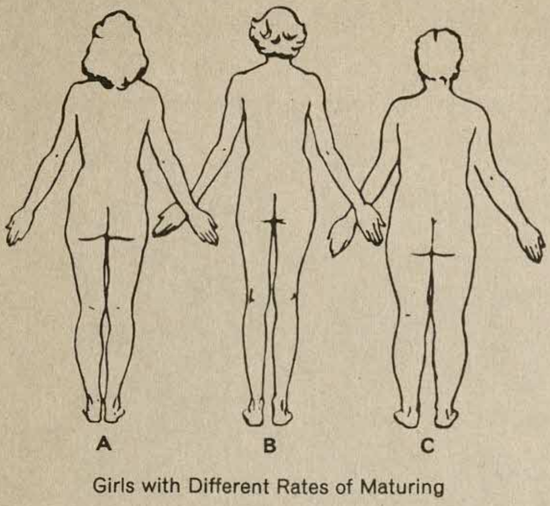

Girls Maturing at Average, Accelerated, and Retarded Rates

up a steadily diminishing part of the total; the legs grow relatively longer to 15 years, and then the stem length increases slightly; the whole body is broad at 15 months, then slenders down, to broaden again later, mostly in the shoulder girdle. The same patterns hold for the girl, except that as she approaches maturity, it is her hips rather than her shoulders that become wide.

In Figure 2 the conventional curve of height for age is compared with a proposed new set of curves. This chart is based on data from several longitudinal studies at the Institute of Child Welfare, primarily the Berkeley Growth Study (Jones & Bayley, 1941) and Macfarlane's Guidance Study (1938). The average curve of height for chronological age of these children is represented by the heavy center line, as "smoothed" by the segments of dotted line. Actually children do not grow in height according to such means of height for age. Their growth is more like that of the solid line that changes direction more abruptly. In this chart the solid curves are based on data that are grouped, not by chronological age, but by their degree of physical maturity. Maturity was determined from readings of the X-rays of the bones by standards that are based on per cent of mature height. The mean age and height were determined for each successive grade of maturity. The curves on either side of the mean represent the trends of growth of children who are one *SD* or more accelerated (the curve on the left) or retarded (on the right) in their skeletal maturing. The broken lines represent the actual growth of fairly extreme cases of a tall accelerated and a short retarded child. Height curves for individual children tend actually to follow or to parallel one or the other of these lines. Figure 3 gives the same data for girls. It is notable that, except for the extremes, the mature average heights of the three groups maturing at different rates are very similar.

Body build tends to be related to rates of

Fig. 4: *Three girls, at 18 years, who matured differently. A was accelerated; B was retarded; C's growth was irregular.*


A
B
C
Girls with Different Rates of Maturing

maturing (Bayley, 1943). That is, the physically accelerated children are more likely to be broad-built, while the retarded children are more often slender. . . .

In Figure 4 are shown 3 girls at maturity. They had different patterns of growth: "A" matured early, "B" was slow in attaining her final height, "C" had a very atypical rate of maturing. Their growth curves by age are shown in Figure 5. Slow-maturing B grew tallest, but she was relatively short before 12 years. . . . Girl C was very accelerated from 8 to 11 years, then matured very slowly until at 13 years she fell in line with the average and maintained this status. Figure 6 gives on the left the curves for these same three girls, each in relation to herself. In attaining their own adult heights, A was advanced and B retarded, while C grew atypically. On the right are the curves of their skeletal maturity, as read on our XM standards that were derived from per cent of mature height groupings. According to these curves, we see that each girl's growth is closely related to her skeletal maturing.

It has become evident, from these and other data on individual growth, that efforts to classify children into somatotypes, or other classes of body build, must take into account the age, or stage of their maturity. If constitutional types as exemplified in the ratio of build [W/H^2] are stable, then we should expect the early-maturers to go through the cycles more rapidly, and the late-maturers more slowly than the average. . . .

There is evidence from other anthropometric ratios and from series of photographs that some children's builds definitely change in some respects. Such changes seem reasonable when we consider the factors that control growth. Goldberg, an endocrinologist, has pointed out (1955) that, from conception to about 2 or 3 years of age, growth is determined by the genes, which set the basic pattern. Then the anterior pituitary growth factor, in conjunction with thyroxin, becomes influential in determining growth during the childhood years. This period is essentially one of elongation of the long bones. Between the ages of 8 and 11 the pituitary gonadotropins begin to function more actively, and as androgens and estrogens appear in quantity, rapid growth is initiated. At the same time general physical maturation occurs together with sexual differentiation in physique. These three different and successively dominating controls of growth are somewhat independent of each other. They may vary both in their intensity and in their temporal patterns. Therefore, at the points of change in their relative influence, certain changes in characteristic growth and build may very well take place.

Age Units of Development

So far, we have been considering physical growth and maturing. Can we present the developmental aspects of behavior and intellectual functioning in ways that will be comparable, and will make possible the evaluation of mental-physical relations in development?

One method that has been used frequently is to convert all of the variables into age units. Olson (1942) has probably developed this method most thoroughly. He has assembled age norms for a variety of physical and behavioral data. From these norms, curves of mental age, height age, weight age, dental age, skeletal age, and the like, are drawn for a single child on one chart. . . . The Organismic Age that Olson has suggested is a composite of all the "ages" employed, and is intended to show whether a child is growing, in general, at an average, accelerated or retarded rate. There are, however, limitations to the usefulness of Organismic Age curves. In the first place, if one looks over a number of such charts, it becomes evident that there is great individuality in the patterns of growth, with considerable independence among the different "ages." For this reason it seems unwise to lump them together into a single Organismic Age. We may, on the other hand, learn quite a bit from studying the spread of a child's different age curves, to see where he is more likely to have trouble as a deviant, and to consider how best he can be helped to fit into a group of his age peers.

But, in the second place, we have no assurance that being advanced or retarded by a year in skeletal age means the same thing as a year's deviation in mental age. If the standard deviations of developmental age for the different variables are not similar, then the direct comparison of "ages" is misleading. Furthermore, one cannot tell whether a child is (say) advanced in weight age because he is naturally large, or because he is accelerated skeletally, or because he is obese. By some study, and comparisons of the child's differences in weight age, height age, and skeletal age, the reasons for the deviations may be deduced. But to do this requires considerable understanding of the factors involved, and they may often be overlooked.

The Quotient and the Standard Score As Measures of Relative Status

If we give up age units of growth, and turn

Fig. 5: *Curves of height of girls A, B, and C. Age at menarche for each is indicated by an* x.

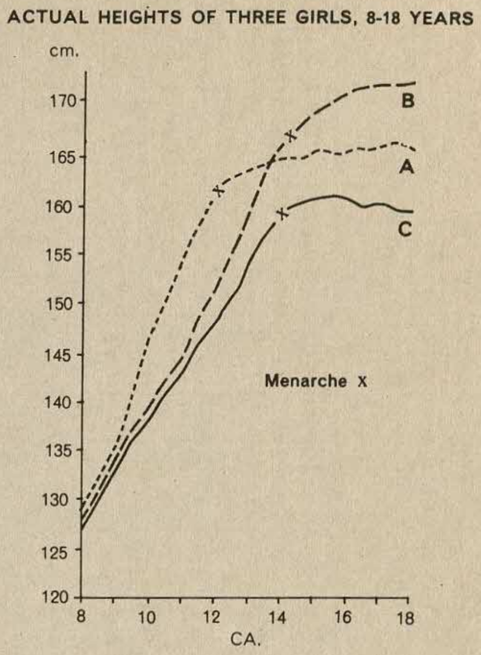


Fig. 6: *Growth curves of girls A, B, and C, showing on the left the per cent achieved by each girl of her own adult stature, for ages 8 to 18; and on the right degrees of skeletal maturity expressed in terms of per cent of mature stature.*

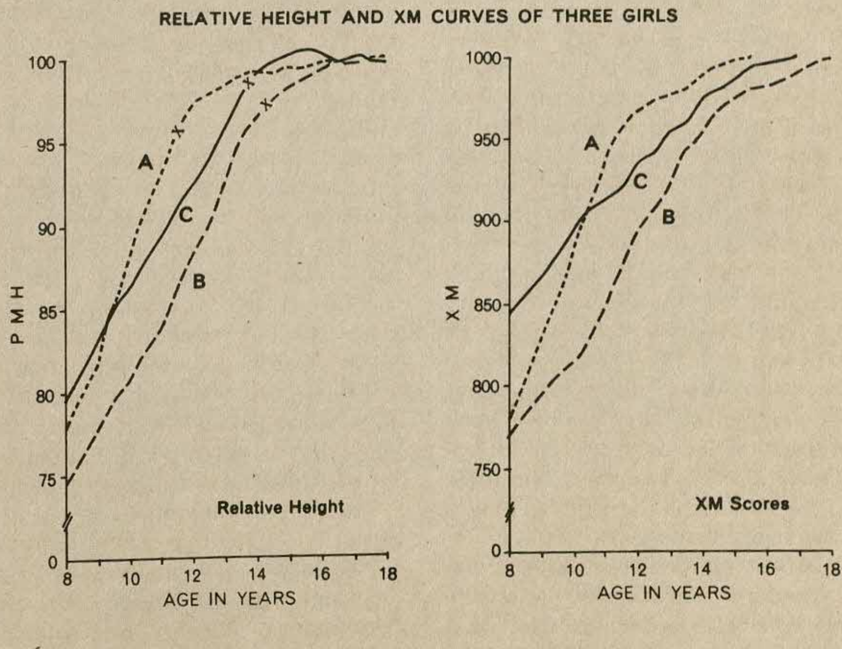
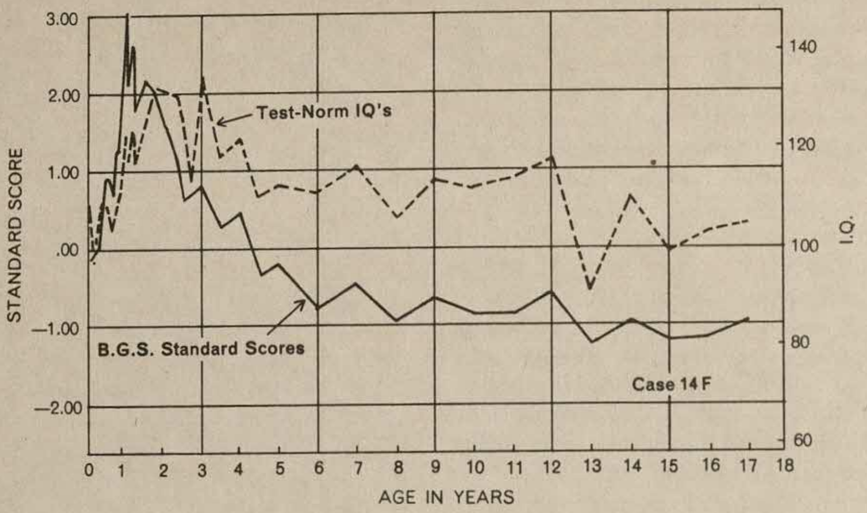


Fig. 7: *Intelligence score curves of Case 14F, expressed as IQs (broken line) and as standard scores according to the Berkeley Growth Study norms.*



to indices of relative status for successive ages, there are two methods that have been used: the ratio or quotient of a given developmental "age" to chronological age (of which the IQ is the best known and most used) and the standard score (or sigma score) based on the means and *SD*s of the normative sample at successive ages. The use of the ratio or quotient has the same limitations as the developmental ages on which it is based. That is, unless all of the age variables under consideration have the same standard deviation the quotients will not be comparable. Furthermore, all of the *SD*s must increase in the same constant way in order to maintain constant quotients for all ages. This is, however, not the case (Bayley, 1949). Both mental and physical measurements, while showing increasing variability with age, generally, have periods in which the *SD*s are expanded and then contracted again.

It would seem that the way out of the dilemma is to use standard scores. In most respects, standard scores avoid the faults of the age quotient. In Figure 7, for example, are curves showing the relative mental scores of one girl from one month to 17 years (Bayley, 1949). The solid line gives her standard scores for the Berkeley Growth Study norms; the broken line gives her IQs. At 12 months she was the brightest child in the study, and her standard scores at this period are higher than her IQs. This is because the variability in scores in the group was very small around one year. Figure 8 gives the IQ and standard score curves for a boy, whose IQ rose to a very high point at 10 years (Bayley, 1949). His standard scores after 4 years are far more stable than his IQs.

From our growth records, we have computed standard scores for several different behavioral and physical measures. Among these are height, weight, head circumference, skeletal maturity, mental and motor development. Standard score patterns for six variables on one girl are shown in Figure 9. A detailed reading of the chart shows that this girl was small and slender, with a small head; she matured slowly and became a tall slender adult; she was consistently superior in intelligence, but more so at some ages than at others; her scores in motor tests varied between superior and average. Around one year of age, when her intelligence score was highest, her pattern of motor scores appears to be almost exactly reversed. In Figure 10, case 13F presents a different pattern, with a tendency for all scores to converge toward the average as she approaches maturity. If space permitted, other cases, with equally diversified curves, could be shown.

. . . These individual cases display great independence in the patterns of growth of the different variables. In some instances there appears to be concomitant variation between those variables that are highly intercorrelated, such as height and weight. Also, there is evidence that

Fig. 8: *Intelligence score curves of Case 5M, expressed as IQs and as standard scores.*

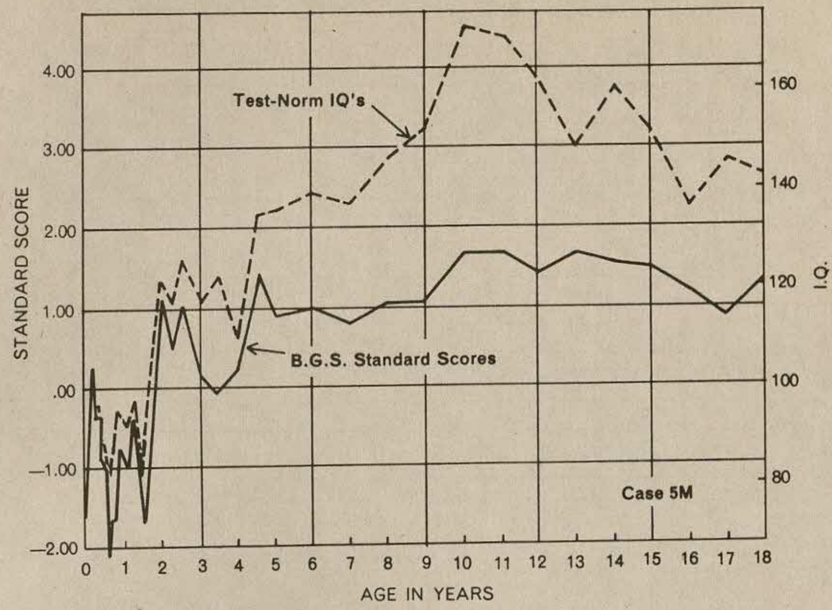
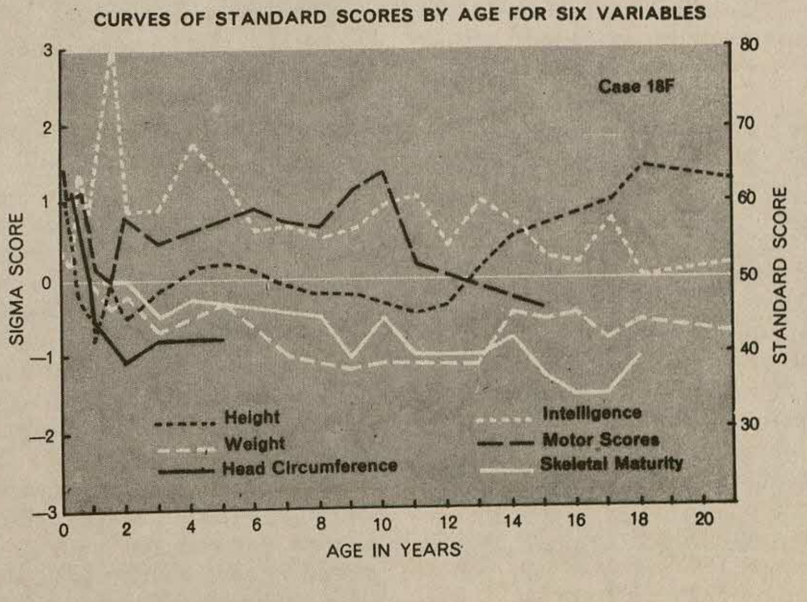


Fig. 9: *Curves of standard scores by age for six variables, for Case 18F.*



certain ones, such as mental and motor development, become more independent of each other with age.

There is one serious drawback in the use of standard scores for curves of this kind, as applied to the first year or two of life. As you may have noticed from the foregoing illustrations, the variation in standard scores during the early years

Fig. 10: *Curves of standard scores by age for six variables, for Case 13F.*

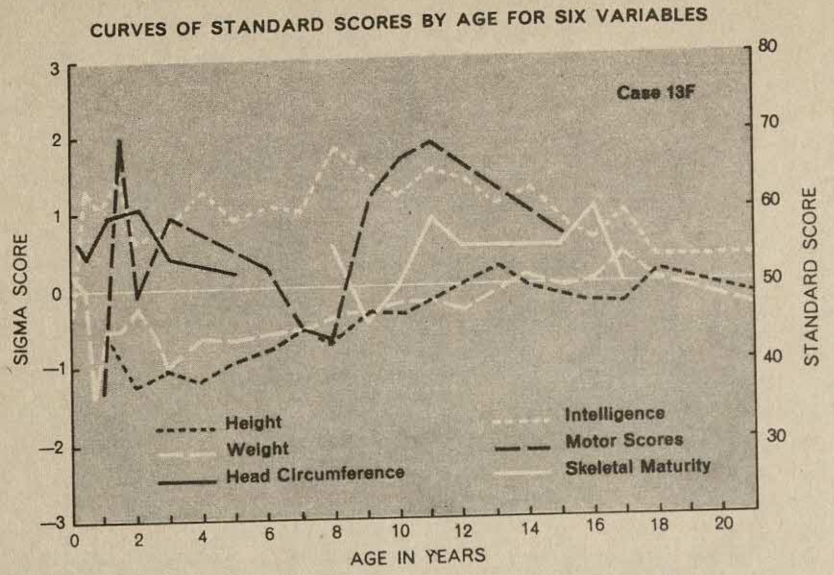
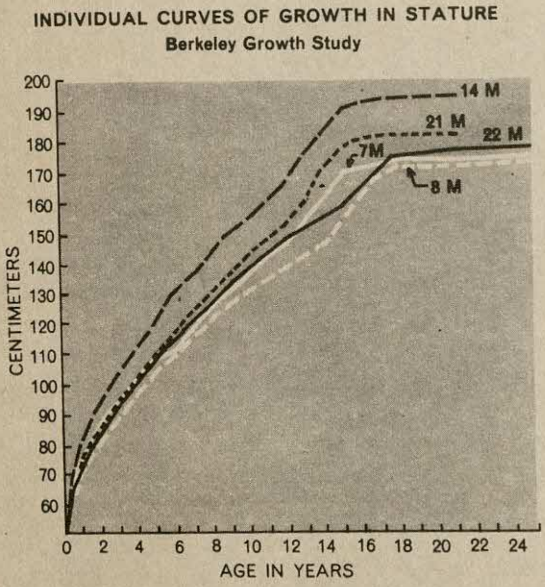


Fig. 11: *Curves of stature by age for five boys in the Berkeley Growth Study.*



is exaggerated. Because the spread of scores at these ages is actually very small as compared with the later ages, a difference of a few centimeters, or of one or two items in the mental test, can mean a difference of half a standard deviation or more. In comparison with the early rapid growth, individual differences at a given age are small, and the infant's growth is much more stable than these wide swings in standard scores would imply.

Absolute Units of Growth

The relative importance of absolute units in evaluating growth and change may be seen in the next series of illustrations. Figure 11 shows the curves for height of five boys, who were selected from the Berkeley Growth Study as the tallest and shortest, at maturity, together with three of intermediate height. The heights of all five are very similar at first, as compared with later ages. Note, also, that there is some crossing of the curves later, because of differences in rates of maturing. Each individual curve is very stable, following a definite pattern. But differences in rates of maturing can make a child's standard score for height shift up and down considerably during the period of puberal growth. Figure 12 shows similar data for five girls. . . .

What about intelligence? Can we show whether or not it presents a similar trend toward increasing diversity with growth? Here we are confronted with the problem of setting up comparable units of increment for intelligence. There is ample evidence that mental ages in normal samples tend to become more variable with age. In working with the Berkeley Growth Study scores on intelligence tests the problem was further complicated by the need to compare scores from several different tests, having different methods of scoring. This has been done by the device of relating all scores to the mean and *SD* earned by this group on the Wechsler-Bellevue scale at 16 years (Bayley, 1955). By adjusting the means and *SD*s earned on the Stanford-Binet and the Terman-McNemar tests at adjacent ages, all scores for the three scales and for all ages, were converted into units of deviation from the 16-year mean, in terms of the 16-year *SD*. This has been called tentatively, the 16-D score. Figure 13 gives the means and *SD*s of these 16-D scores from one month to 21 years. They are not "absolute" units, but they do give a general picture of growth relative to the status of this group at 16 years. Again the variability in scores for the first 2 years is very restricted as compared with the later ages.

In Figure 14 are individual curves for 5 boys selected to represent wide differences in adult intelligence. Note that all five have been tested through 25 years, and in all of them the scores are still increasing. These curves, too, are less regular than the height curves, but perhaps no less regular than the weight curves. One gets the impression both of differences in rates of maturing and of differences in inherent capacity. Superimposed on these general trends are the fluctuations that may be related to differences in motivation or drive, or to other factors such as emotional distractions, or differences in the content of the tests used at different ages. Figure 15 shows the same thing for 5 girls. When we see how closely alike the scores are in the first two years, it seems obvious that this is at least one reason why it is impossible to predict later intelligence from scores in infancy.

Individual Differences in Rates of Maturing

It must be remembered that some of the changes in patterns of standard scores are due to the fact that children differ in their time schedule in going through the process of growing up. . . .

Fig. 12: *Curves of stature by age for five girls in the Berkeley Growth Study.*

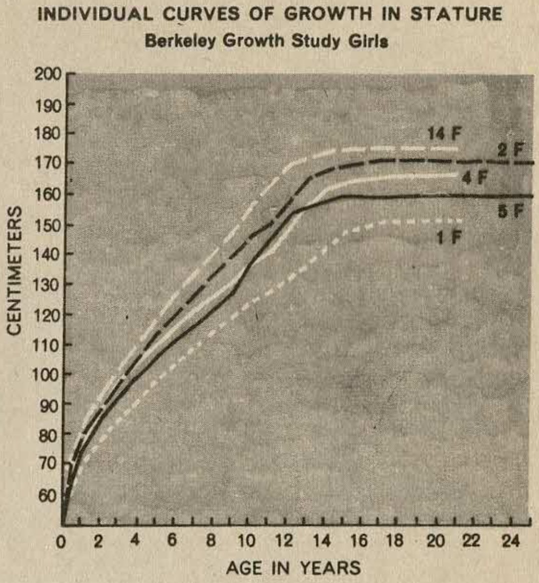
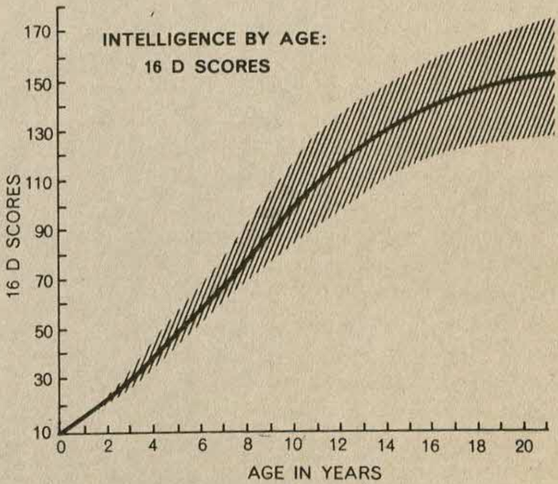


Fig. 13: *Curve showing the mean and* SD *of intelligence scores of the Berkeley Growth Study children, from birth to 21 years. Scores from several different tests were converted into units based on the mean and* SD *of the 16-year test scores.*

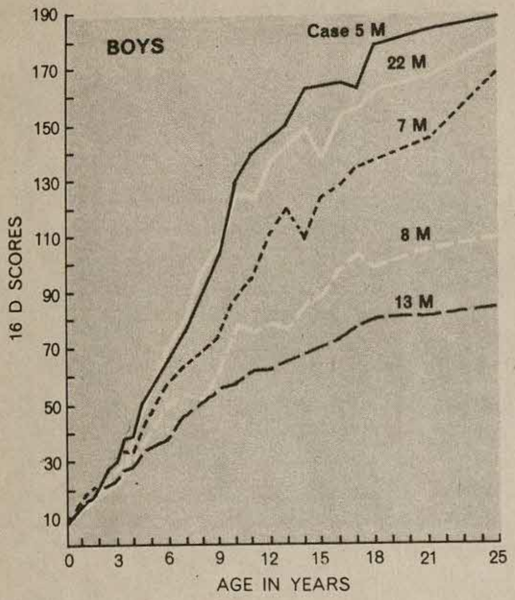


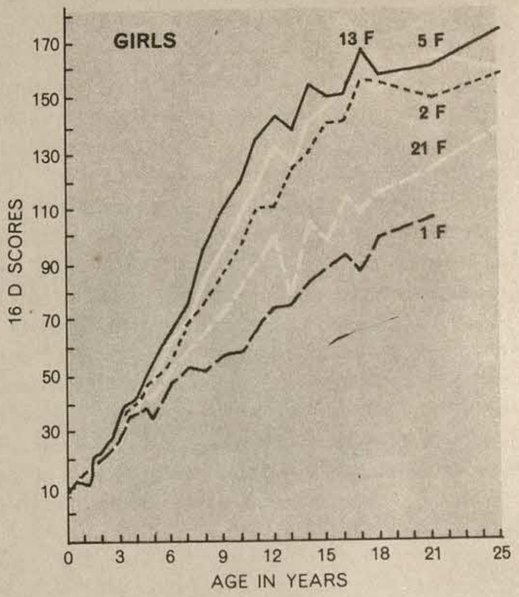
Is there some way to control or rule out individual differences in size or ability, and thus simplify the study of differences in rates of growth? As I mentioned earlier, when we have available a series of measures taken during growth, together with a stable measurement at maturity, then that stable measure can be used

Fig. 14: *Individual curves of intelligence of 5 boys, from 1 month to 25 years.*

Fig. 15: *Individual curves of intelligence of 5 girls.*

INDIVIDUAL CURVES OF 16 D SCORES (INTELLIGENCE)





as a point of reference, against which any child's other measures can be evaluated. For example, we can take each child's height at the time the epiphyses of his long bones fuse and he has practically stopped growing, and call this figure 100 per cent (Bayley & Pinneau, 1952). The portion of this stature attained at each earlier measuring can then be computed. In this way all of the children's rates of growth toward their final statures can be compared without the confusing differences in size. . . .

We may try to approximate this condition for the intelligence test scores, but there are several differences between the two. As noted already, the units of intelligence are not directly comparable, and can be considered only as approximations of more and less ability. But there is another difficulty. These children did not conveniently stop improving in their intelligence test scores. Fourteen out of 15 so far tested at 25 years are still gaining. For present purposes, however, it was necessary to choose a point for comparison. Twenty-one years seemed most reasonable, as the age at which recent investigations have usually found the high point in intelligence test scores. Using 21 years, then, as 100 per cent, and the 16-D scores, each child's earlier per cent of adult intelligence was computed. Figure 16 gives the curves of the means for the per cents of adult status for height and for intelligence, and for boys and girls separately of the Berkeley Growth Study Sample. It is evident that growth in the two factors differs in several ways. We find obvious sex differences in one and not in the other. The periods of rapid and slow growth do not coincide, but seem almost to alternate. A study of annual increments indicates that intelligence is accelerating most rapidly up to 5 or 6 years, with another short spurt between 9 and 11. Height, on the other hand, though most rapid during the first 2 years, grows at a declining rate until 8 or 10, with adolescent spurts between 10 and 12 for the girls and 11 and 15 for the boys.

With such independence in the rates of growth in these variables we should not expect to find very closely related mental and physical maturing in the same child. A few correlation coefficients bear on this point. These children's heights and 16-D intelligence scores are positively correlated, as seen by the *rs* in Figure 17, connected by the solid lines. Actually, the *rs*, which vary around .40, are surprisingly high, when compared with similar correlations reported for other studies. Shown here also, connected by dotted lines, are the correlations between skeletal maturation and the intelligence scores. These *rs* tend to be positive, also, until skeletal maturity is reached.

When, however, we compare the per cent of mature height with per cent of 21-year intelligence scores, as seen in the two lower curves, the picture is very different. These *rs* are not large, but they are, with only one exception, of negative sign. In this small group of children, it seems that the more intelligent children tend to be both larger than average, and accelerated in physical maturing. But within the individual child his own mental and physical rates of growth are not concomitant. If anything, there is a suggestion that those who are slower in physical maturing approach their 21-year intelligence sooner. That is, if we think of it in the more usual way, there may be a tendency for the less able (and these are also, generally the socio-economically less favored children) to be

Fig. 16: *A comparison of growth in stature and intelligence, when both are expressed as fractions of mature status.*

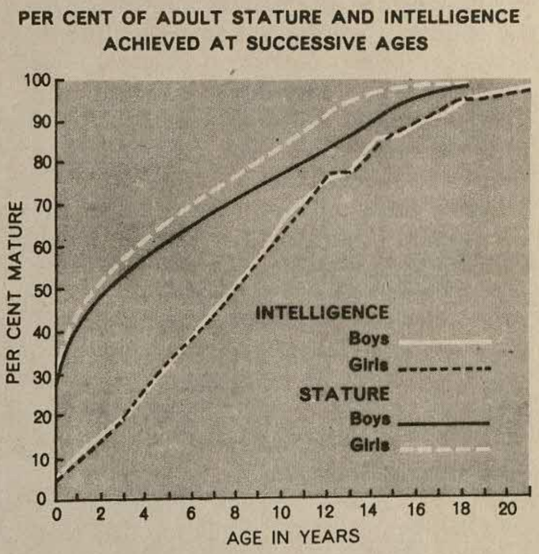
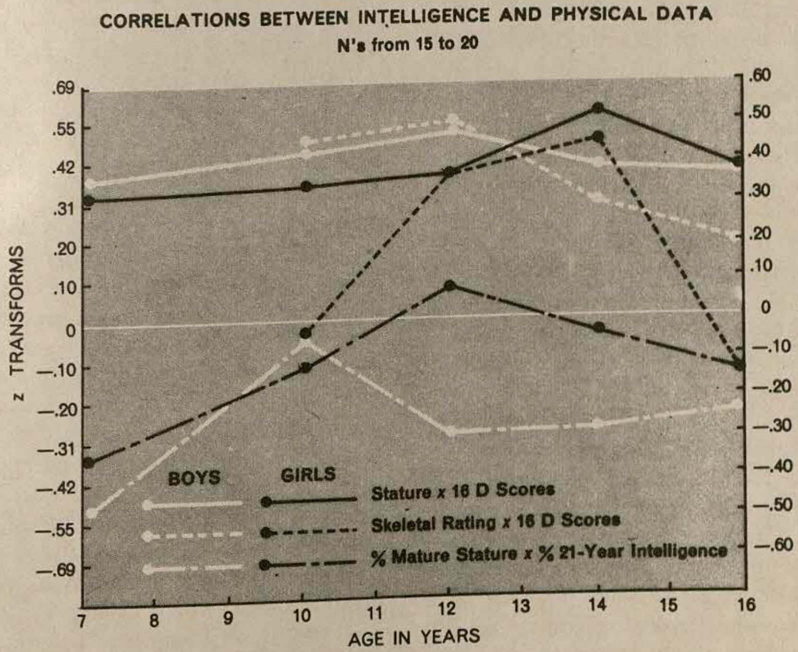


Fig. 17: *Correlations for boys and girls separately, of intelligence scores with stature and skeletal age, and of per cent of 21-year intelligence score with per cent of mature stature.*



slow in maturing physically. But they attain most of their intellectual growth early, and slow down in this respect as they approach adulthood. These same children, as I have reported earlier (1933), are the ones who in infancy tended to make high scores on the mental tests. That is, the children's mental test scores for the first 6 months correlate negatively with their scores at 3 years or later. There is also a negative correlation between the scores at ages 3 to 7 months with the amount of their parents' education (Bayley & Jones, 1937).

In this presentation, I have neglected the environmental factors, as they bear on individual patterns in development. Other areas not dealt with are emotions and various aspects of personality, which cannot easily be fitted into development sequences or scales. These factors have been omitted, in part, because it seems to me necessary first to understand the conditions under which they must be evaluated if we are to estimate their effectiveness.

It is evident that the whole process has become very complex. But, keeping in mind as many as possible of the relevant factors just reviewed, we can go down the list of cases, and try to decide which array of factors is pertinent to the course of each child's physical and mental development. Let me list briefly a few examples that will illustrate a few kinds of relationships we have found.

In a pair of fraternal twins, who were born just a little prematurely, there were closely similar patterns of rapid early growth in both size and behavioral functions. At this very early stage, structure and function were not yet independent. However, later, their mental and physical curves became very dissimilar: One twin became heavy, the other remained slender. One twin with real problems of social adjustment made significantly higher scores on the intelligence tests than the better adjusted twin. Possibly the very difficulties of adjustment caused Twin A to turn to intellectual preoccupations as a form of compensation.

A boy who never seemed highly motivated, intellectually, earned fair intelligence scores, but had difficulty graduating from high school. He suddenly evidenced strong motivation on his 25-year test, with a large increase in score. A short time before this last test he had struck out on his own for the first time, on a job that was not connected with the family business, and he was belatedly making good.

A girl made phenomenal intellectual growth between 2 and 10 years. During much of this time she was invalided with asthma. She was strongly competitive with an older sibling, and used her enforced rest periods in reading and imaginative games.

A boy who was very slow in both motor and mental development during infancy made rapid intellectual progress after two years: He was highly verbal, but never gained much skill in his motor coordinations. Here the emerging independence of intellectual capacities seems to have permitted their more rapid growth.

One boy was undernourished in infancy, with accompanying retardation in physical maturing and growth, but this seemed unrelated to his relatively steady mental development.

Several children suddenly slowed down in their skeletal maturing, at around 13 years, with subsequent prolonged slow pre-adolescent growth. Two of these, with their later spurts of growth, eventually attained taller statures than had been predicted before 13. In these cases there appeared to be more or less normal delays in starting the steroid phase of growth.

There was at times evidence that a child was "bored" with repetition of the same old mental tests, and the resulting poor effort lowered the score. An adolescent girl, suddenly aware of her plumpness and more curvaceous figure, would diet and lose weight. And of course, severe illnesses often caused temporary slowing down of the physical growth processes.

There are as many more such examples as there are children in the Study. However, there is not space here for them all or even to go into the details to support these brief descriptive statements. But the data are there, and it can be done. The more one goes over these cases, the more evident it becomes that the pattern of growth in each child is unique. We can compare him with his peers, and with his own past history. Often, after the fact, we can find an explanation of the causes of his deviations, some of which seem to be inherent and some environmental. With the accumulation of this kind of information, we can hope gradually to develop general rules that will permit us to make some evaluations and some tentative predictions, and even, in some instances, to correct undesirable deviations.

I hope these evidences of the complexity in the processes of child development will not be too discouraging. On the contrary, they should spur us on to learn more about their intricate interrelationships. Only in this way may we eventually set up optimal conditions for growth and development.

REFERENCES

BAYLEY, NANCY. Mental growth during the first three years. *Genet. Psychol. Monogr.*, 1933, *14*.

BAYLEY, NANCY. Size and body build of adolescents in relation to rate of skeletal maturing. *Child Develpm.*, 1943, *14*, 47–90.

BAYLEY, NANCY. Consistency and variability in the growth of intelligence from birth to eighteen years. *J. genet. Psychol.*, 1949, *75*, 165–196.

BAYLEY, NANCY. On the growth of intelligence. *Amer. Psychologist*, 1955, *10*, 805–818.

BAYLEY, NANCY, & JONES, H. E. Environmental correlates of mental and motor development: A cumulative study from infancy to six years. *Child Develpm.*, 1937, *8*, 329–341.

BAYLEY, NANCY, & PINNEAU, S. A. Tables for predicting adult height from skeletal age: Revised for use with the Gruelich-Pyle hand standards. *J. Pediat.*, 1952, *40*, 423–441.

GOLDBERG, MINNIE B. What makes us grow as we do. *J. Amer. med. Women's Ass.*, 1955, *10*, 110–116.

JONES, H. E., & BAYLEY, NANCY. The Berkeley Growth Study. *Child Develpm.*, 1941, *12*, 167–173.

MAC FARLANE, J. W. Studies in child guidance: I. Methodology of data collection and organization. *Monogr. Soc. Res. Child Develpm.*, 1938, *3*, No. 6.

OLSON, W. C., & HUGHES, B. O. The concept of organismic age. *J. educ. Res.*, 1942, *35*, 525–527.

5

Stability of Achievement and Recognition Seeking Behaviors from Early Childhood Through Adulthood

Howard A. Moss and Jerome Kagan

A major question facing the developmental psychologist is if, and in what way, an individual's behavior patterns, motives, and attitudes persist over time. Bayley's article pointed out patterns of physical and intellectual growth; this selection from the Fels Research Institute demonstrates that there are certain consistencies of behavior from childhood through adulthood. Like Bayley, Moss and Kagan have used longitudinal research, and it is only through such careful research that the uniqueness of the individual's personal growth can be analyzed.

The supposition that selected adult response patterns are established at an early age is a primary assumption of developmental theory. Although literary documents and psychotherapy protocols have provided anecdotal support for this hypothesis, more objective validation has been difficult to obtain. The present paper is a second report that has emerged from a larger project on the stability of childhood behavior. The first paper indicated that dependent behavior in girls showed moderately high stability from the early school years through young adulthood (Kagan & Moss, 1960). The present report is concerned with the developmental consistency of two related behaviors: the tendency to strive for *(a)* mastery of selected skills (achievement behavior), and *(b)* social recognition through acquistion of specific goals or behaviors (recognition behavior).

The achievement variable emphasizes mastery of intellectual, athletic, mechanical, and artistic skills as well as competence in specialized crafts. Social recognition is obtained through acquisition of most of the above behaviors. For intellectual competence, athletic ability, acquisition of money, and positions of power in social groups are the primary methods of obtaining social recognition in the cultural milieu of our middle class population. Thus, the overt behaviors involved in achievement and recognition strivings overlap to some degree.

In an attempt to differentiate between these two variables, the investigators evaluated the degree to which the individual's mastery behavior was directed at satisfaction of an internal standard of excellence in order to gain self-approval (achievement motivation), in contrast to seeking approval from the social environment (recognition motivation). This is a difficult differentiation to make. The data to be presented reveal a high, positive correlation between ratings of these two behavioral variables. This interdependence suggests that it may be impossible to measure the "desire to improve at a skill" inde-

---

From *Journal of Abnormal and Social Psychology*, 1961, *62*, 504–513

pendent of the individual's "desire for social recognition" for this improvement.

METHOD

Subjects and General Procedure

The subjects were 36 males and 35 females from the Fels Research Institute's longitudinal population. They were enrolled in the project at birth, during the years 1929–1939. At the time of a recent adult assessment (1957–1959) they were between 20 and 29 years of age. The subjects came from predominantly middle class backgrounds; over half of the group were married; 70% had college degrees or were enrolled in a college, and the majority were living within a 30-mile radius of the institute. The adult group included 55 Protestants, 15 Catholics, and 1 Jew.

The heart of this study consists of correlations between the childhood information on these subjects and their adult behavior. The childhood data included *(a)* longitudinal observations of the child's behavior during the first 14 years of life in a variety of settings, *(b)* observations of the mother-child interaction during these years, *(c)* TAT protocols obtained in adolescence, and *(d)* annual Stanford-Binet intelligence test scores during the ages 5–11. Although the data collected during adulthood (age range 20–29) sampled a variety of techniques, this report utilizes only two sources of adult information, 5 hours of interview, and a TAT protocol.

Longitudinal Observations: Birth to Age 14

As a standard procedure of the Fels longitudinal program, psychologists or psychologically trained personnel summarized their observations of the child in the home, in the Fels nursery school and day camp, and in the subject's public school. The home reports were based on a visit to the home where mother and child were observed for half-day sessions. These home visits were generally made semiannually for the first 6 years of life and annually from 6 to 12. Most of the mothers were interviewed each year for the first 14 years of the child's life. The nursery school summaries were based on semiannual, free-play sessions from age 2.5 to 5. The sessions usually consisted of 15 consecutive half-day periods in groups of 10–12 children. Day camp typically consisted of an annual 2 week session of half-day periods during age 6–10 in which free and structured group activities were observed. Public school visits, made semiannually, consisted of a half-day observation of the child in his routine classroom activities. Finally, the subjects of age 6–14 were interviewed each year at the institute and a summary of the interview was prepared. All of the longitudinal reports for each subject were collated in chronological order and placed in the subject's individual file.

Scoring of longitudinal variables. A comprehensive list of rating scale variables (seven-point scale) was defined for the purpose of evaluating the narrative material just outlined. The material for each subject was divided into four age periods: 0–3, 3–6, 6–10, and 10–14. The senior author, who had no knowledge of the adult psychological status of the subjects, first read all the information for each subject for age 0–3 and made those ratings for which he had adequate information. Following a period of interpolated work, he studied each subject's material for age 3–6 and again made his ratings. This procedure was repeated for ages 6–10 and 10–14. A period of approximately 6 months intervened between the evaluation of the data for any one subject for each age period. This paper deals only with the stability of achievement and recognition behaviors and abridged definitions of these variables follow.

Childhood Variables

Achievement behavior: (Rated for ages 0–3, 3–6, and 6–10). This variable assessed the degree to which the subject tended to persist with challenging tasks, games, and problems, and his involvement in activities in which a standard of excellence was applicable. For 0–3, emphasis was given to persistence with perceptual-motor activities (e.g., making block towers, stringing beads, drawing, and coloring). For ages 3–6 and 6–10 the greatest weight was given to interest in and persistence with intellectual, mechanical, athletic, and fine motor activities.

For age 10–14 the general achievement variable defined above was differentiated into three variables dealing with different achievement areas (intellectual, mechanical, and athletic).

Intellectual achievement: (Rated for age 10–14). This variable assessed the degree to which the subject attempted to master language and numerical skills and showed involvement in the acquisition of knowledge.

Mechanical achievement: (Rated for age 10–14).

This variable assessed the degree to which the subject attempted to master mechanical skills and manifested involvement in activities such as carpentry, construction of model vehicles, engines and motors, and craft work.

Athletic achievement: (Rated for age 10–14). This variable assessed the degree to which the subject attempted to master and showed involvement in athletic activities. These behaviors included swimming, hiking, baseball, football, basketball, tennis, acrobatics, and track events.

Recognition seeking behavior: (Rated for ages 6–10 and 10–14). This variable assessed the subject's striving to obtain goals that led to recognition from parents, teachers, and peers. The behaviors emphasized in the rating were *(a)* grades in school and school honors, *(b)* stated desire for status-laden vocations or ostentatious material goods, *(c)* striving for leadership in teams or clubs, *(d)* attempts to get recognition from farm activities (e.g., raise the best calf, the highest corn, etc.).

Maternal Variables

Maternal acceleration of developmental skills in child: (Rated for ages 0–3, 3–6, and 6–10). The home visits and maternal interviews yielded information on the mother's behavior and attitudes toward her child. The maternal variable that is directly relevant to the subject's achievement behavior was called *maternal acceleration*. It was defined in terms of the degree to which the mother showed concern over the subject's cognitive and motor development, and the degree to which she exhibited desires for precocious achievement in her child. The rating reflected the degree to which the mother "pushed" the subject's development beyond his abilities and her concern with his general achievement level.

Adult Interview

The junior author, who had no knowledge of the subjects' childhood information, interviewed each subject and rated him (seven-point scale) on a variety of variables. The definitions of the variables related to achievement and recognition seeking behaviors follow.

Achievement behavior. This variable evaluated the subject's behavioral attempts to master tasks for which "self-satisfaction" rather than social recognition was the *salient* goal. In achievement behavior, the subject was striving to attain a *self-imposed* standard of excellence. The rating was based on the subject's emphasis and concern with task mastery in his job and avocational pursuits.

Recognition seeking behavior. This variable evaluated the subject's behavioral attempt to obtain symbols of status and social recognition. The rating was based on evidences of strivings for *(a)* vocational recognition, *(b)* academic awards and honors, *(c)* positions of leadership or recognition in community or vocational groups, *(d)* concern with conspicuous material display, *(e)* striving for upward mobility in social class position.

Concern with intellectual competence. This variable assessed the value the subject placed upon intelligence, knowledge, academic achievement, and intellectual superiority regardless of whether the goal was to satisfy inner standards or to obtain social recognition.

Reliability of Longitudinal and Adult Interview Ratings

A random sample of 32 tape recorded adult interviews were independently studied and rated by a second judge to assess the reliability of the junior author's adult ratings. The reliabilities of the longitudinal variables were also assessed through independent ratings, by a second judge, of samples of 50–60 cases at each of the four age periods. The reliabilities of the adult and child ratings were determined by product-moment correlation coefficients. For the adult ratings of achievement behavior, recognition behavior, and intellectual concern the reliability coefficients were .84, .99, and .98, respectively. With the exception of one child behavior variable, the reliabilities of the longitudinal ratings ranged from .74 to .90 with a median coefficient of .81. The one low longitudinal reliability was for child's achievement for age 0–3 ($r = +.35$; $p < .01$; two-tailed).

TAT Achievement Fantasy: Adolescent and Adult Protocols

Early adolescent (median age of 14–6) protocols were available for 67 of the 71 subjects, and all 71 subjects were administered TAT stimuli following the adult interview. The adolescent protocol was based on seven cards from the Murray (1943) series (Cards 1, 5, 14, 17BM, 3BM, 6BM, and 3GF). The male adult protocol was based on 13 cards (4, 8BM, 7BM, 6BM, 12M, 17BM, 13MF, 14, 3BM, 5, 1, 3GF, and 18GF). The adult females were also administered 13 cards (4, 6GF, 12F, 2, 8GF, 17BM, 13MF, 14, 3BM, 5, 1, 3GF, and 18GF). For both the adolescent and adult protocols achievement themes were scored according to the scheme described by McClelland, Atkinson, Clark, and Lowell (1953). Since incidence of the subcategories of the McClelland scoring system were infrequent, only stories in which achievement behavior was the major aspect of the plot were considered. These are scored Ach Th in the McClelland scheme. For the adolescent protocol, there was a lack of comparability among the examiners with respect to the inquiry questions and only the spontaneous verbalization of the subject was scored. Agreement between two independent coders was 95%. The longitudinal and interview ratings of achievement and recognition behavior were made *without knowledge* of the subject's adolescent or adult TAT stories. Thus, the behavior and interview ratings were independent of each other and of the TAT thematic scores.

IQ Change

Each child was given the Stanford-Binet, Forms L and M alternately, annually from ages 5 through 11 by the same psychologist. The mean IQ for the entire Fels population is about 120 (*SD* of 15). For each subject, a smoothed plot of his IQ scores was obtained by averaging his three IQ scores around each age. For example, a child's smoothed or average IQ at age 6 was the result of averaging his IQ scores at ages 5, 6, and 7; his smoothed IQ at age 10 was the average of his IQs at ages 9, 10, and 11. This procedure tends to remove the chance variation associated with any one IQ score and has been used in other studies (Kagan, Sontag, Baker, & Nelson, 1958; Sontag, Baker, & Nelson, 1958). Each subject's smoothed IQ at age 6 was then subtracted from his smoothed IQ at age 10 and the resulting difference was used as a measure of IQ change. As with achievement themes, the child and adult achievement ratings were made without knowledge of the subject's IQ or his IQ change score.

In summary, four independent sources of data were analyzed; child and maternal behaviors for the first 14 years of life, adult behavior, adolescent and adult achievement themes, and childhood IQ change scores.

Statistical Analysis

Relationships among the following variables were evaluated: *(a)* childhood achievement and maternal acceleration ratings with the adult interview ratings, *(b)* adolescent achievement themes with adult achievement themes, *(c)* adolescent and adult achievement themes with the longitudinal and adult ratings, and *(d)* IQ change scores with the childhood and adult ratings. Product-moment correlations were used except when the TAT achievement score was involved. Since achievement themes were not normally distributed, contingency coefficients were used for all tests of association using this variable. Mechanical achievement for age 10–14 was the only variable for which there was a significant sex difference; the boys having a higher mean rating than the girls ($p < .05$; two-tailed).

RESULTS

Stability of Achievement and Recognition Behaviors

Table 1 presents the relationships between the child and adult ratings of achievement and recognition behavior, as well as the relation between maternal acceleration and the adult achievement variables. There are several important results in this table. The rating of achievement behavior for age 6–10 showed a significant, positive association with all three adult variables for both sexes. The rating of achievement for age 3–6 was predictive of adult behavior for the females but not for the males, a finding that suggests the earlier emergence of stable achievement strivings in girls' development than in boys. Of the three achievement behaviors rated for age 10–14, only intellectual mastery was predictive of adult achievement for both sexes. Involvement in mechanical activities was predictive of adult achievement for boys but not for girls. Athletic achievement showed no relationship to the rating of general adult achievement, and was negatively associated with intellectual concern for adult males ($p < .02$).

Recognition seeking behavior for age 6–10 was also predictive of adult achievement behavior. A few of the child variables were moderately intercorrelated and the three adult variables were highly intercorrelated (Tables 4, 5, and 6). This lack of independence makes some of the stability correlations between childhood and adulthood somewhat redundant.

Maternal Acceleration and Adult Behavior

Maternal concern with the child's developmental progress during the first 10 years of life showed low to moderate correlations with adult achievement behavior. The maternal rating for age 6–10 was not a better predictor of adult behavior than the maternal rating for the first 3 years of life. Moreover, the age 0–3 rating was associated with all three adult, achievement variables for girls, while it predicted only recognition behavior for adult males.

Stability of TAT Achievement Fantasy

Although different sets of TAT pictures were used in obtaining the adolescent and adult protocols, the three pictures that usually elicited achievement stories were presented at both administrations. Cards 1, 14, and 17BM, which elicited 77% of all the achievement themes, were common to both protocols. The strong tendency for these particular cards to elicit achievement themes has been noted in another study (Kagan & Moss, 1959). A typical achievement theme to Card 1 concerned a boy who wanted to master the violin and/or become a famous violinist. A typical achievement story to Card 17BM involved a person who was in a rope climbing contest and wanted to do his best to win. A common achievement story to Card 14 concerned an artist or student who had been working hard and was looking forward to fame and success as a result of his accomplishments.

The stability of the TAT achievement score between the adolescent and adult protocols was determined through the use of contingency coefficients. The stability coefficients were .34, .36, and .31 for boys, girls, and total group ($p < .10$, $< .05$, $< .02$; one-tailed). Thus, achievement themes also showed some degree of stability over this 10-year period. These data extend the findings of an earlier investigation (Kagan & Moss, 1959), in which the authors reported a 3 year stability coefficient of .32 ($p < .01$) for achievement themes obtained at median ages of 8–9 and 11–6. The stability coefficients between the adolescent and adult protocols are of the same magnitude as those found for the earlier age period.

Validity of Achievement Themes: Relations with Child and Adult Behavior

Contingency coefficients were computed relating the occurrence of adolescent and adult

Table 1: *Relation between longitudinal ratings of childhood achievement and early maternal acceleration with adult achievement behavior (product-moment correlations)*

Childhood Variables	Age	Adult Variables: Achievement, Males	Achievement, Females	Recognition, Males	Recognition, Females	Intellectual Concerns, Males	Intellectual Concerns, Females
Recognition	6–10	.47***	.40**	.42**	.48***	.37**	.55****
	10–14	.25	.20	.36*	.39**	.24	.40**
Achievement	0–3	–.12	–.02	.01	–.22	–.08	–.02
	3–6	–.03	.45**	–.11	.49***	.13	.44**
	6–10	.46***	.38**	.57****	.51***	.69****	.49***
Achievement							
Intellectual	10–14	.40**	.42**	.60****	.56***	.66****	.49***
Mechanical	10–14	.20	.20	.46**	.02	.47**	.27
Athletic	10–14	–.18	.01	–.17	–.09	–.47**	.02
Maternal Acceleration	0–3	.22	.36*	.44*	.41*	.09	.36*
	3–6	.31	.09	.24	.12	.42**	.12
	6–10	.14	.33*	.16	.23	.32*	.43**

* $p < .05$; *one-tailed.*
** $p < .02$; *one-tailed.*
*** $p < .01$; *one-tailed.*
**** $p < .001$; *one-tailed.*

Table 2: *Relation between TAT achievement themes and child and adult achievement behavior* (*contingency coefficients*)

Longitudinal Variable	Age	Adolescent TAT (Median age 4-6)		Adult TAT (Median age 25)	
		Males	Females	Males	Females
Recognition	6–10	.21	.17	.26	.22
	10–14	–.20	–.18	–.31	.39
Achievement	0–3	–.15	–.25	–.20	.16
	3–6	.42*	.19	.19	.36
	6–10	.24	.15	.13	.30
Achievement					
Intellectual	10–14	.30	–.25	.26	.16
Mechanical	10–14	.31	–.62***	.63****	–.50**
Athletic	10–14	–.20	.12	.12	.17
Maternal Acceleration	0–3	.11	.51**	.25	–.08
	3–6	.37	.23	.27	.28
	6–10	.51***	.26	.24	.41*
Adult interview variables (median age 25)					
Recognition		.17	.25	.40**	.52****
Achievement		.19	.44***	.37*	.52****
Intellectual Concerns		.44***	.25	.31	.59****

* *$p < .05$; one-tailed.*
** *$p < .02$; one-tailed.*
*** *$p < .01$; one-tailed.*
**** *$p < .001$; one-tailed.*

achievement themes with the longitudinal and adult achievement ratings. These results are presented in Table 2. The highest and most consistent relations were between the adult achievement themes and adult interview ratings. The only significant relation between adult themes and the childhood ratings held for mechanical achievement ($C = +.63$; $p < .001$ for boys, and $-.50$; $p < .02$ for girls).

The adolescent TAT was also more predictive of adult behavior than it was of the childhood ratings. Adolescent achievement themes predicted adult achievement behavior for women ($C = +.44$; $p < .01$) and intellectual concerns for men ($C = +.44$; $p < .01$). Adolescent achievement themes showed minimal association with the child's achievement behavior. The only significant positive association was with age 3–6 achievement for boys. Once again the rating of mechanical achievement for girls was negatively associated with achievement themes. This negative correlation may be due to the fact that this is the only variable for which markedly different behavioral referents were used in rating the two sexes. For boys, involvement in carpentry, engines, motors, and model airplanes was emphasized in the rating. These activities are sex-typed and girls showed no interest in them. Participation in craft work (making jewelry, leather articles) and sewing was also used as evidence of involvement in mechanical activities and girls tended to choose these behaviors.

Maternal acceleration during the first 10 years of life showed suggestive relationships with the adolescent achievement themes. For example, maternal acceleration for age 0–3 predicted achievement themes at adolescence for girls ($C = +.51$; $p < .02$), but not for boys. Maternal acceleration for age 6–10 predicted adolescent achievement themes for boys ($C = +.51$; $p < .01$) and adult achievement themes for girls ($C = +.41$; $p < .05$).

In summary, the adult and adolescent TAT stories showed moderate correlations with adult achievement but minimal association with the childhood achievement ratings. Maternal acceleration was associated, to some degree, with adolescent achievement themes.

IQ Increase and Achievement Behavior

The difference between the child's smoothed IQ at age 6 and 10 was used as a measure of IQ change. Earlier studies (Kagan *et al.*, 1958; Sontag *et al.*, 1958) have demonstrated that the amount of increase in IQ correlated both with independent behavioral indices of achievement strivings for age 6–10, and with early adolescent (age 10–14) achievement stories. These latter relations remain significant when the influence of the child's IQ at age 6 is statistically controlled. The present data allowed for a partial validation of these results and an extension of the Sontag et al. findings for adolescence and adulthood.

Table 3 presents the correlations between changes in IQ during age 6–10 and the longitudinal and adult behaviors. The amount of IQ increase was a fairly sensitive predictor of both intellectual achievement for age 10–14 ($r = .37$ and .41 for boys and girls; $p < .01$), and concern with intellectual competence in adulthood ($r = .49$ and .42; $p < .01$). These results support and extend the earlier studies and indicate that amount of IQ increase during the first 4 years of school is a moderately accurate index of the subject's motivation to master intellectual tasks during adolescence and early adulthood. It is important to note that IQ change showed no relation to mechanical or athletic strivings for boys, and was negatively associated with athletic achievement for girls ($r = -.46$; $p < .02$). Thus, IQ increase is not a general measure of achievement strivings for all areas of task mastery. The IQ change measure predicts all three adult achievement ratings because the three adult variables are heavily weighted with concern over intellectual competence. Finally, the maternal acceleration rating for age 6–10 showed a positive relation with IQ change for girls ($r = +.54$; $p < .01$) but not for boys.

Intercorrelations among the Measures

There were, as might be anticipated, positive correlations among the achievement and recognition ratings. Tables 4, 5, and 6 present the intercorrelations among the variables that were rated for ages 6–10, 10–14, and adulthood. The correlations for the males are above and to the right of the diagonal; the female data are to the left and below the diagonal.

For all three age periods there were high, positive correlations among the achievement and recognition variables. For age 10–14, recognition behavior was highly correlated with achievement strivings in the intellectual area, but only minimally related to mechanical or athletic achievement. This finding suggests that, for this middle class sample, mastery of intellectual

Table 3: *Relation between IQ change and childhood and adult achievement variables (product-moment correlations)*

Longitudinal Variables	Age	Males	Females
Recognition	6–10	.24	.21
Recognition	10–14	.41**	.09
Achievement	0–3	.13	.04
Achievement	3–6	−.02	.24
Achievement	6–10	.39***	.47***
Achievement-intellectual	10–14	.37*	.41**
Achievement-mechanical	10–14	.15	.14
Achievement-athletic	10–14	−.16	−.46**
Maternal acceleration	0–3	−.06	.20
Maternal acceleration	3–6	−.03	−.12
Maternal acceleration	6–10	.10	.54***
Adult interview variables			
Recognition		.48***	.25
Achievement		.38***	.38**
Intellectual concern		.49***	.42***

** $p < .05$; one-tailed.*
*** $p < .02$; one-tailed.*
**** $p < .01$; one-tailed.*

Table 4: *Intercorrelations among variables rated for age 6-10*

	Recognition	Achievement	Maternal Acceleration
Recognition	—	.77***	.57***
Achievement	.60***	—	.59***
Maternal acceleration	.39*	.44***	—

Note.—Data for males are in upper right; for females, in lower left.
** $p < .05$; two-tailed.*
**** $p < .01$; two-tailed.*

Table 5: *Intercorrelations among variables rated for age 10-14*

	Recognition	Intellectual Achievement	Mechanical Achievement	Athletic Achievement
Recognition	—	.74***	.23	.04
Intellectual achievement	.60***	—	.53***	−.11
Mechanical achievement	.17	.32	—	−.23
Athletic achievement	.24	.12	.07	—

Note.—Data for males are in upper right; for females, in lower left.
**** $p < .01$; two-tailed.*

Table 6: *Intercorrelations among adult interview variables*

	Achievement	Recognition	Intellectual
Achievement	—	.72***	.73***
Recognition	.79***	—	.72***
Intellectual	.77***	.84***	—

Note.—Data for males are in upper right; for females, in lower left.
**** $p < .01$; two-tailed.*

skills is the primary method chosen to obtain social recognition. Perhaps for lower class samples this generalization might be less valid. The high correlations between recognition behavior and intellectual concern in adulthood, together with the fact that maternal acceleration predicted both variables, suggests that it is difficult to separate "recognition seeking behavior" from "attempts to improve intellectual competence."

DISCUSSION

Stability of Achievement Strivings

The results indicate that strivings for intellectual mastery are moderately stable from the school years through early adulthood. This behavioral disposition emerges as a stable phenomenon at ages 3–6 for girls and 6–10 for boys. The stability of the behavior ratings is paralleled by the moderate stability of TAT achievement stories over a shorter age span. Moreover, achievement stories in adolescence and adulthood also predicted the adult behavior ratings. This consistent cluster of correlations adds construct validity to the TAT achievement variable and support to the conclusion that this class of behaviors is stable over time.

Involvement in athletics for age 10–14 showed no strong, positive relation to either IQ increase or adult achievement behavior and, in a few instances, negative relationships occurred. This was not because the interviewer failed to assess adult involvement in this particular activity. Rather, many of the adults who had been involved in athletics as early adolescents were not overly concerned with task mastery as adults and they tended to avoid intellectual activities.

The majority of the sample regarded positions of responsibility, intellectual challenge, and knowledge of the environment as highly desirable goals. If a subject had strong achievement motives he tended to gratify them through intellectually oriented endeavors. It is suggested that the mass media and social environment differentially emphasize the importance of different skills in accordance with the sex and age role characteristics of the individual. For adults, there tends to be an emphasis on intellectual competence and a de-emphasis on active mastery of athletic skills. Moreover, intellectual mastery is less involved in potential sex role conflict than mechanical or athletic behaviors. To excel at sports is one of the defining characteristics of masculinity. Some boys become involved in athletics in order to maintain their sex role identity and avoid peer rejection. An athletic girl will be subject to peer rejection for excessive participation in athletics. Thus, athletic mastery is under the control of motives and conflicts related to sex role identification in addition to needs for task mastery.

This latter point raises the question of the appropriate definition of achievement behavior and motivation. It is suggested that the concept of a general achievement motive is too broad a term, and it may be useful to replace this construct with a series of variables that relate to more specific behaviors. It seems more reasonable to talk about "desire to improve intellectual skills," or "desire to improve athletic skills" than to use the more global concept of need achievement. Individuals strive to perfect skills in different areas, and the motivations for these strivings are multiple. Prediction and comprehension of these phenomena might be facilitated if there was some differentiation among the behaviors and motives that are involved in task mastery.

The lack of predictive power of age 0–3 mastery behavior might have been due to the greater difficulty in rating this variable (the interrater reliability was .35). On the other hand, the behavioral referents for this rating differed from those used to assess mastery for the older age periods. Since 2-year-olds do not initiate intellectual or athletic mastery behavior, persistence with simple, perceptual-motor tasks (stringing beads, building towers) was the basis for this early rating. A high rating for 0–3 reflected a high threshold for satiation with simple, sensorimotor activities. At the older ages, the achievement rating was based on involvement with problem solving behaviors that were more similar in form to adult achievement behavior. The age 0–3 rating is dynamically different from the symbolic behaviors that characterize achievement during the preschool years.[1] This statement is supported by the fact that achievement for age 0–3 was negatively correlated with achievement for age 3–6 ($r = -.20$) and age 6–10 ($r = -.03$), and showed no relationship to achievement themes or IQ change. Persistence with simple sensorimotor tasks during the first 2 or 3 years of life is not an index of future intellectual, achievement strivings. The 2-year-old who will sit for 20 minutes trying to put a peg in a hole is not necessarily the ambitious scholar of the fifth grade.

Maternal Acceleration and Achievement

The ratings of maternal concern with the child's developmental skills were heavily weighted with encouragement of intellectual progress. The most consistent correlates of maternal acceleration were found with the ratings of adult concern with intellectual competence. Maternal acceleration for age 0–3 was slightly more predictive of adult behavior for girls than for boys. Similarly, maternal acceleration for age 6–10 was more predictive of IQ increase for daughters than for sons. The sex difference between these latter two correlations was significant at the .05 level. It is suggested that since the girl was more likely than the boy to identify with the mother, maternal encouragement of intellectual mastery should have had a greater effect on the development of the girl than on the boy.[2]

TAT Achievement Stories

Achievement themes on the TAT were moderately stable and were correlated with adult achievement behavior. The fact that the correlations were as high for females as for males, although the three critical cards illustrated male heroes, raises some question concerning the validity of the hero hypothesis. Since Cards 1, 14, and 17BM all picture a male in a potential achievement situation, one might expect that achievement themes for women would not be highly correlated with their achievement behavior. The present results indicate that the production of achievement themes may be more influenced by the subject's conception of what behaviors are appropriate for the hero, than by the degree of identification of storyteller with hero. Perhaps high achievement girls conceptualize the male role as being more associated with task mastery than do low achievement girls.

Atkinson (1957) has suggested that achievement themes have differential validity depending on whether or not the criterion task engages the subject's motivation. The achievement variables used in this study (ratings of overt behavior, IQ increase scores) measured "real life" behaviors that would be expected to engage the subject's motivation. The positive correlations obtained indicate that achievement themes are valid indices of intellective mastery when the conditions under which the behavioral samples are obtained are motive arousing.

Limitations on Generalizability

Although the stability correlations for achievement behavior are fairly high, the nature of this particular sample favored stability. The social milieu of these subjects remain constant throughout the first 17 years of their lives, and the parents and peers of these subjects retained their same values. The degree of stability obtained with this sample might not hold for populations that were more mobile, for different ethnic or social class groups, or for children subjected to major developmental traumata.

Social Reinforcement and Stability

The stability of achievement behavior is congruent with general reinforcement theory. Each time achievement strivings are rewarded through social approval or internal feelings of satisfaction, the strength of this behavioral tendency should be increased. If achievement strivings lead to failure, these behaviors should extinguish. The child who attains scholastic honors through effort is rewarded by the social environment, and this experience frequently leads to an expectancy of future success for similar behavior. This rewarding experience, coupled with the strong cultural approval for intellectual competence, increases the probability that the child will continue to engage in intellectual tasks. On the other hand, persistent failures in intellectually challenging situations are likely to lead to an expectancy of failure, and these expectancies can result in avoidance and/or withdrawal from involvement in intellectual behavior.

---

[1] The differences in the content of mastery behaviors for ages 0–3 and 3–6 are analogous to Piaget's (1952 description of intellectual development. Piaget suggests that during the first 2 years the child is in the sensorimotor stage of intelligence in which simple perceptual-motor activity is salient. During the subsequent preschool years the child's intellectual activity becomes more symbolic and more comparable to adult problem-solving behavior.

[2] Research in progress at the institute, under the direction of Vaughn J. Crandall and Walter Katkovsky, is assessing the role of both mother and father in adoption of achievement motives and behaviors.

SUMMARY

This paper summarized results from a larger investigation on the stability of behavior from childhood through adulthood. This investigation dealt specifically with the long term stability of achievement and recognition seeking behaviors in subjects who were part of the Fels Research Institute's longitudinal population.

The subjects were 36 males and 35 females for whom extensive longitudinal information was available from birth through 14 years of age. One psychologist, who had no knowledge of the adult behavior of these subjects, studied narrative reports based on observations of the child in a variety of settings, and rated each child on achievement and recognition seeking behaviors for four age periods: 0–3, 3–6, 6–10, and 10–14. In addition, ratings were made for maternal acceleration of developmental skills for the first three age periods. A second psychologist, who had no knowledge of the childhood information, interviewed each subject in adulthood (age range 20–29) and rated him on three variables related to achievement and recognition seeking behavior in adulthood.

In addition, the following information was available for most subjects: *(a)* a 7 card TAT protocol administered during early adolescence (median age 14-6) and a 13 card TAT protocol following the adult interview, and *(b)* annual Stanford-Binet IQ tests from ages 5 through 11 which furnished an IQ change score over the years 6–10.

The major results were as follows:

1. Both achievement and recognition striving behaviors for age 6–10 showed significant positive correlations with similar behaviors during adulthood, the correlations ranged from .38 to .68 and all were significant at the .05 level or better.

2. Involvement in intellectual activities for age 10–14 showed high positive correlations with achievement behavior in adulthood while involvement in athletics during these years showed no positive relationship with adult achievement behavior. Involvement in mechanical tasks for age 10–14 showed positive correlations with adult achievement behavior for boys but not for girls.

3. Maternal acceleration of the child's developmental skills during the first 3 years of life predicted adult achievement behavior for women but not for men. Maternal acceleration of developmental skills during age 6–10 showed moderate correlations with adult concern with intellectual competence in both sexes.

4. Achievement stories told during adolescence and adulthood showed high positive correlations with adult achievement behavior. Maternal acceleration of developmental skills showed suggestive correlations with the occurrence of achievement stories in both adolescence and adulthood.

5. The amount of increase in IQ score during the years 6–10 showed high positive correlations with the ratings of achievement behavior during adulthood. In addition, increase in IQ showed positive correlations with age 6–10 achievement behavior, and with strivings for intellectual competence during age 10–14. Amount of IQ increase showed negative correlations with athletic achievement for age 10–14.

The results suggested that achievement strivings during the first 4 years of school are a moderately good index of future achievement behavior during adolescence and adulthood. There were high correlations between strivings for social recognition and intellectual achievement, and it was suggested that these variables are intimately related in a middle class population.

REFERENCES

ATKINSON, J. W. Motivational determinants of risk-taking behavior. *Psychol. Rev.*, 1957, *64,* 359-372.

KAGAN, J., & MOSS, H. A. The stability and validity of achievement fantasy. *J. abnorm. soc. Psychol.*, 1959, *58,* 357-364.

KAGAN, J., & MOSS, H. A. The stability of passive and dependent behavior from childhood through adulthood. *Child Develpm.*, 1960, *31,* 577-591.

KAGAN, J., SONTAG, L. W., BAKER, C. T., & NELSON, VIRGINIA L. Personality and IQ change. *J. abnorm. soc. Psychol.*, 1958, *26,* 261-266.

McCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

MURRAY, H. A. *Thematic Apperception Test Manual.*

Cambridge, Mass.: Harvard University Press, 1943.

PIAGET, J. *The origins of intelligence in children.* New York: International Universities Press, 1952.

SONTAG, L. W., BAKER, C. T., & NELSON, VIRGINIA L. Mental growth and personality development. *Monogr. Soc. Res. Child Develpm.*, 1958, *23*, No. 68.

6

The Genetic Approach to the Psychology of Thought

Jean Piaget

Jean Piaget's work, which spans over thirty years, was neglected for a long time by American psychologists because of his research methodology and his interest in cognition. With the current renaissance of interest in cognition, however, Piaget's work has become the source of much research effort both in Europe and in the United States—particularly in the development of new curricula in education. Although no one short paper adequately represents his system, this article gives a brief introduction.

From a developmental point of view, the essential in the act of thinking is not contemplation —that is to say, that which the Greeks called "theorema"—but the action of the dynamics.

Taking into consideration all that is known, one can distinguish two principal aspects:

1. The formal viewpoint which deals with the configuration of the state of things to know—for instance, most perceptions, mental images, imageries.
2. The *dynamic* aspect, which deals with transformations—for instance, to disconnect a motor in order to understand its functioning, to dissociate and vary the components of a physical phenomenon, to understand its causalities, to isolate the elements of a geometrical figure in order to investigate its properties, etc.

The study of the development of thought shows that the dynamic aspect is at the same time more difficult to attain and more important, because only transformations make us understand the state of things. For instance: when a child of 4 to 6 years transfers a liquid from a large and low glass into a narrow and higher glass, he believes in general that the quantity of the liquid has increased, because he is limited to comparing the initial state (low level) to the final state (high level) without concerning himself with the transformation. Toward 7 or 8 years of age, on the other hand, a child discovers the preservation of the liquid, because he will think in terms of transformation. He will say that nothing has been taken away and nothing added, and, if the level of the liquid rises, this is due to a loss of width, etc.

The formal aspect of thought makes way, therefore, more and more in the course of the development to its dynamic aspect, until such time when only transformation gives an understanding of things. To think means, above all, to understand; and to understand means to arrive at the transformations, which furnish the reason for the state of things. All development of thought is resumed in the following manner: a construction of operations which stem from actions and a gradual subordination of formal aspects into dynamic aspects.

The operation, properly speaking, which constitutes the terminal point of this evolution is, therefore, to be conceived as an internalized action reversible (example: addition and subtraction, etc.) bound to other operations, which form with it a structured whole and which is characterized by well defined laws of totality (example: the groups, the lattice, etc.). Dynamic totalities are clearly different from the "gestalt" because those are characterized by their nonadditive composition, consequently irreversible.

So defined, the dynamics intervene in the construction of all thought processes; in the structure of forms and classifications, of relations and serialization of correspondences, of numbers, of space and time, of the causality, etc. One could think at first glance that space and geometry add to the formal aspect of thought. In this way one conceived of the geometric science in the past, considering it impure mathematics, but applicable to perception and intuition. Modern geometry, since *Le Programme d'Erlangen* by F. Klein, has tended, like all other precise disciplines, to subordinate the formal to the dynamic. The geometries are, indeed, understood today as

---

From *Journal of Educational Psychology*, 1961, *52*: 275–281.

relying all on groups of transformation, so that one can go from one to the other by characterizing one less general "subgroup" as part of a more inclusive group. Thus geometry too rests on a system of dynamics.

Any action of thought consists of combining thought operations and integrating the objects to be understood into systems of dynamic transformation. The psychological criteria of this is the appearance of the notion of conservation or "invariants of groups." Before speech, at the purely sensory-motor stage of a child from 0 to 18 months, it is possible to observe actions which show evidence of such tendencies. For instance: From 4–5 to 18 months, the baby constructs his first invariant, which is the schema of the permanent object (to recover an object which escaped from the field of perception). He succeeds in this by coordinating the positions and the displacements according to a structure, which can be compared to what the geometricians call "group displacements."

When, with the beginning of the symbolic function (language, symbolic play, imagerie, etc.), the representation through thought becomes possible, it is at first a question of reconstructing in thought what the action is already able to realize. The actions actually do not become transformed immediately into operations, and one has to wait until about 7 to 8 years for the child to reach a functioning level. During this preoperative period the child, therefore, only arrives at incomplete structures characterized by a lack in the notion of combinations and, consequently, by a lack of logic (in transitivity, etc.).

In the realm of causality one can especially observe these diverse forms of precausality, which we have previously described in detail. It is true that a certain number of authors—Anglo-Saxon above all—have severely criticized these conclusions, while others have recognized the same facts as we have (animism, etc.). Yet, in an important recent book (which will appear soon) two Canadian authors, M. Laurendeau and A. Pinard, have taken the whole problem up once again by means of thorough statistics. In the main points they have come to a remarkable verification of our views, explaining, moreover, the methodological reasons for the divergencies among the preceding authors.

At about 7 to 8 years the child arrives at his first complete dynamic structures (classes, relations, and numbers), which, however, still remain concrete—in other words, only at the time of a handling of objects (material manipulation or, when possible, directly imagined). It is not before the age of 11 to 12 years or more that operations can be applied to pure hypotheses. At this latter level, a logic of propositions helps complete the concrete structures. This enlarges the structures considerably until their disposition.

The fundamental genetic problem of the psychology of thought is hence to explain the formation of these dynamic structures.

Practically, one would have to rely on three principal factors in order to explain the facts of development: maturation, physical experience, and social interaction. But in this particular case none of these three suffice to furnish us with the desired explanations—not even the three together.

Maturation. First of all, none of these dynamic structures are innate, but they form very gradually. (For example: The transitivity of equalities is acquired at approximately 6½ to 7 years, and the ability of linear measure comes about only at 9 years, as does the full understanding of weights, etc.) But progressive construction does not seem to depend on maturation, because the achievements hardly correspond to a particular age. Only the order of succession is constant. However, one witnesses innumerable accelerations or retardations for reasons of education (cultural) or acquired experience. Certainly one cannot deny the inevitable role which maturation plays, but is determined above all by existing possibility (or limitation). They still remain to be actualized, which brings about other factors. In addition, in the domain of thought, the factors of innateness seem above all limitative. We do not have, for example, an intuition of space in the fourth dimension; nevertheless we can deduce it.

Physical Experience. Experiencing of objects plays, naturally, a very important role in the establishment of dynamic structures, because the operations originate from actions and the actions bear upon the object. This role manifests itself right from the beginning of sensory-motor explorations, preceding language, and it affirms itself continually in the course of manipulations and activities which are appropriate to the antecedent stages. Necessary as the role of experience may be, it does not sufficiently describe the construction of the dynamic structures —and this for the following three reasons.

First, there exist ideas which cannot possibly be derived from the child's experience—for instance, when one changes the shape of a small ball of clay. The child will declare, at 7 to 8 years, that the quantity of the matter is conserved. It does so before discovering the conservation of weight (9 to 10 years) and that of volume (10 to 11 years). What is the quantity of a matter independently of its weight and its volume? This is an abstract notion corresponding to the "substance" of the pre-Socratic physicists. This notion is neither possible to be perceived nor measurable. It is, therefore, the product of a dynamic deduction and not part of an experience. (The problem would not be solved either by presenting the quantity in the form of a bar of chocolate to be eaten.)

Secondly, the various investigations into the learning of logical structure, which we were able to make at our International Center of Genetic Epistemology, lead to a very unanimous result:[1] one does not "learn" a logical structure as one learns to discover any physical law. For instance, it is easy to bring about the learning of the conservation of weight because of its physical character, but it is difficult to obtain the one of the transitivity of the relationship of the weight:

$$A = C \text{ if } A = B \text{ and } B = C$$

or the one of the relationship of inclusion, etc. The reason for this is that in order to arrive at the learning of a logical structure, one has to build on another more elementary logical (or prelogical) structure. And such structures consequently never stem from experience alone, but suppose always a coordinating activity of the subject.

Thirdly, there exist two types of experiences:

1. The physical experiences show the objects as they are, and the knowledge of them leads to the abstraction directly from the object (example: to discover that a more voluminous matter is more or less heavy than a less voluminous matter).

2. The logicomathematical experience supposes to interrelate by action individual facts into the world of objects, but this refers to the result of these actions rather than to the objects themselves. These interrelations are arrived at by process of abstractions from the actions and their coordinates. For instance, to discover that 10 stones in a line always add up to 10, whether they are counted from left to right or from right to left. Because then the order and the total sum have been presented. The new knowledge consists simply in the discovery that the action of adding a sum is independent of the action of putting them in order. Thus the logicomathematical experience does not stem from the same type of learning as that of the physical experience, but rather from an equilibration of the scheme of actions, as we will see.

Social Interaction. The educative and social transmission (linguistic, etc.) plays, naturally, an evident role in the formation of dynamic structures, but this factor does not suffice either to entirely explain its development, and this for two reasons:

First, a certain number of structures do not lend themselves to teaching and are prior to all teaching. One can cite, as an example, most concepts of conservation, of which, in general, the pedagogs agree that they are not problematic to the child.

The second, more fundamental, reason is that in order to understand the adult and his language, the child needs means of assimilation which are formed through structures preliminary to the social transmission itself—for instance, an ancient experience has shown us that French-speaking children understand very early the expression *"quelques unes de mes fleurs"* [some of my flowers] in contrast to *"toutes mes fleurs"* [all my flowers], and this occurs when they have not yet constructed the relation of inclusion:

$$\text{Some A are part of all B; therefore } A < B$$

In conclusion, it is not exaggerated to maintain that the basic factors invoked before in order to explain mental development do not suffice to explain the formation of the dynamic structures. Though all three of them certainly play a necessary role, they do not constitute in themselves sufficient reason and one has to add to them a fourth factor, which we shall try to describe now.

This fourth factor seems to us to consist of a general progression of equilibration. This factor intervenes, as is to be expected, in the interaction of the preceding factors. Indeed, if the development depends, on one hand, on internal factors (maturation), and on the other hand on external factors (physical or social), it is self-evident that these internal and external factors equilibrate each other. The question is then to know if we are dealing here only with momentary compromises (unstable equilibrium) or if, on the contrary, this equilibrium becomes more and more stable. This shows that all exchange (mental as well as biological) between the organisms and the milieu (physical and social) as composed of two poles: *(a)* of the *assimilation* of the given external to the previous internal structures, and *(b)* of the *accommodation* of these structures to the given ones. The equilibrium between the assimilation and the accommodation is proportionately more stable than the assimilative structures which are better differentiated and coordinated.

---

[1] See *Etudes d'Epistomologie Genetique*, Vols. 7 and 10

It is this equilibrium between the assimilation and accommodation that seems to explain to us the functioning of the reversible operations. This occurs, for instance, in the realm of notions of conservation where the invariants of groups do not account for the maturation and the physical experience, nor for the sociolingual transmission. In fact, dynamic reversibility is a compensatory system of which the idea of conservation constitutes precisely the result. The equilibrium (between the assimilation and the accommodation) is to be defined as a compensation of exterior disturbances through activities of the subject orientated in the contrary direction of these disturbances. This leads us directly to the reversibility.

Notice that we do not conceive of the idea of equilibrium in the same manner as the "gestalt theory" does, which makes great use of this idea too, but in the sense of an automatical physical equilibrium. We believe, on the contrary, that the mental equilibrium and even the biological one presumes an activity of the subject, or of the organism. It consists in a sort of matching, orientated towards compensation—with even some overcompensation—resulting from strategies of precaution. One knows, for instance, that the homeostasis does not always lead to an exact balance. But it often leads to overcompensation, in response to exterior disturbances. Such is the case in nearly all occurrences except precisely in the case of occurrences of a superior order, which are the operations of reversible intelligence, the reversible logic of which is characterized by a complete and exact compensation (inverted operation).

The idea of equilibrium is so close to the one of reversibility that G. Brunner, in a friendly criticism of one of our latest books appearing in the *British Journal of Psychology,* proposes to renounce the idea of equilibrium because the notion of the reversibility seems sufficient to him. We hesitate to accept this suggestion for the following three reasons:

First, reversibility is a logical idea, while the equilibrium is a causal idea which permits the explanation of reforms by a means of a probabilistic shema. For instance, in order to explain the formation of the idea of conservation, one can distinguish a certain number of successive stages, of which each is characterized by the "strategy" of a progress of compensation. Now it is possible to show[2] that the first of these strategies (only bearing upon one dimension, to the neglect of others) is the most probable at the point of departure, and further, that the second of these strategies (with the emphasis on a second dimension) *becomes* the most likely—as a function of the result of the first. And, finally, that the third of these strategies (oscillation between the observed modifications upon the different dimensions and the discovery of their solidarity) *becomes* the most likely in the functioning of the results of the preceding, etc. From such a point of view the process of equilibration is, therefore, characterized by a sequential control with increasing probabilities. It furnishes a beginning for causal explanations of the reversibility and does not duplicate the former idea.

Secondly, the tendency of equilibrium is much broader for the operation than the reversibility as such, which leads us to explain the reversibility through the equilibrium and not the reverse. In effect, it is at this level of the obvious regulations and sensory-motor feedbacks that the process of equilibration starts. This in its higher form becomes intelligence. Logical reversibility is therefore conceivable as an end result and not as a beginning and the entire reversibility follows the laws of a semireversibility of various levels.

Thirdly, the tendency to equilibrate does not only explain this final reversibility, but also certain new synthesis between originally distinct operations. One can cite in this regard an example of great importance: the serial of whole numbers. Russell and Whitehead have tried to explain the basic set of numbers through the idea of equivalent classes, without recourse to the serial order. This means that two classes are believed to be equivalent, if one can put their respective elements into a reciprocal arrangement. Only when this relationship relies on the quality of the objects (an A put into relation with an A, a B with a B, etc.) one does not get the quantity. If this relationship is made exclusive of the qualities (an Individual A or B put into relationship with an Individual B or A) then there exists only one way to distinguish the elements from each other. In order not to forget one, or not to count the same twice, one must deal with them in succession and introduce the serial factor as well as the structure of classes. We may then say, psychologically speaking, that the sequence of whole numbers is synthesis between two groupings qualitatively distinct, the fitting of the classes and serialization, and that this synthesis takes place as soon as one excludes the qualities of the elements in question. But how does this synthesis occur? Precisely by a gradual process of equilibration.

---

[2] *Logique et Equilibre.* Vol. 2 of *Etudes d'Epistomologie Genetique.*

On the one hand the child who develops his ideas from numbers is in possession of structures enabling him to fit them into classes (classifications). But if he wants to be exclusive of qualities in order to answer to the question "how many," be becomes unable to distinguish the elements. The disequilibrium which appears, therefore, obliges the child to resort to the idea of order and take recourse to arranging these elements into a lineal row. On the other hand, if the child arranges the elements as 1, 1, 1, etc., how would he know, for instance, how to distinguish the second from the third? This new disequilibrium brings him back to the idea of classification: The "second" is the element which has but one predecessor, and the "third" is one that has two of them. In short, every new problem provokes a disequilibrium (recognizable through types of dominant errors) the solution of which consists in a re-equilibration, which brings about a new original synthesis of two systems, up to the point of independence.

During the discussion of my theories, Brunner has said that I have called disequilibrium what others describe as motivation. This is perfectly true, but the advantage of this language is to clarify that a cognitive or dynamic structure is never independent of motivational factors. The motivation in return is always solidary to structural (therefore cognitive) determined level. The language of the equilibrium presents that activity, that permits us to reunite into one and the same totality those two aspects of behavior which always have a functional solidarity because there exists no structure (cognition) without an energizer (motivation) and vice versa.

7

The Development of Quantitative Thinking: A Systematic Replication of Piaget's Studies

David Elkind

One of the tasks of research is to replicate the findings of other scientists. Particularly in the field of human development; where so many variables may influence behavior, it is important to retest in order to verify the results. Elkind is currently engaged in a replication of Piaget's work and is thus, in effect, exploring the influence of different cultures on the development of thinking.

PROBLEM

One of the continuing needs in psychology is for the replication and refinement of significant experiments. This is especially true for the experiments of the Swiss psychologist, Jean Piaget. For over 30 years Piaget has been studying the development of children's thinking but many of his studies (Piaget, 1928; 1929; 1930) have been devoid of statistical methods and systematic design. The present replication study takes one of Piaget's later (1952) investigations as its starting point. . . .

Piaget assumes that the child's quantitative thinking develops and that the child's success in comparing quantity earmarks this development. By studying the child's responses to quantity comparisons of increasing difficulty Piaget sought to trace the developmental changes in the form and content of the child's quantitative thought.

In his work Piaget distinguishes three types of perceived quantity by which things can be compared without actual measurement. The simplest type of perceived quantity he speaks of as *gross*. Gross quantities are single perceived relations between objects (longer than, larger than) which are not coordinated with each other. A more complex type of quantity Piaget calls *intensive*. Intensive quantities are perceived quantity relations taken two by two (longer and wider, taller and thicker). The most complex type of perceived quantity Piaget calls *extensive*. Extensive quantities are unit relations between objects

---

From *Journal of Genetic Psychology,* 1961, *98:* 37–46.

(X is half of Y, X is twice Y, etc.). Extensive quantities are logical constructs which must be attained by abstraction or reasoning but Piaget assumes that once attained, the subject perceives them directly as perceptually given properties of the object.

To test for success in comparing quantity Piaget had subjects of various ages compare two amounts of material several times. The two amounts were successively arranged so that sometimes a comparison of gross, sometimes a comparison of intensive, and sometimes a comparison of extensive quantity was the minimum requirement for a successful result. Experimenting with a variety of materials, including liquids, sticks, and beads, Piaget found three, age related, hierarchically ordered stages of success in comparing quantity.

Children at the first stage (usually age 4) succeeded only when a comparison of gross quantity was the minimum requirement for success. At the second stage (usually age 5) children succeeded when a comparison of gross or intensive quantity was the minimum requirement for success. Third stage children (usually age 6–7) succeeded when a comparison of gross, intensive, or extensive quantity was the minimum requirement for a successful result. Piaget reports his findings and interpretations with the aid of a great many illustrative experimental examples but without the aid of statistics.

In the present study the following hypotheses, derived from Piaget's findings, were tested: *(a)* that Age Groups (4, 5, 6–7) differ significantly in their success in comparing quantity; *(b)* that the Type of Quantity required (gross, intensive, or extensive) by the test significantly affects successful comparison; *(c)* that the Type of Material (sticks, liquids, or beads) significantly affects successful comparison; *(d)* that Age and Type of Quantity jointly affect successful comparison (the statistical test of Piaget's stages); *(e)* that Type of Quantity and Type of Material jointly affect successful comparison; *(f)* there is a positive correlation between performance on the three Types of Material; *(g)* there is a positive correlation between success in comparing quantity and performance on the W.I.S.C.

METHOD

Subjects

Eighty school and pre-school children living in Pittsfield, Mass., were tested. They were divided into three age groups and the number in each group was: Age four, 18; Age five, 40; Ages six and seven, 22. For each group the mean age and standard deviation were: Age four, 54.0 and 2.6 mos.; Age five, 64.9 and 2.7 mos.; Ages six and seven, 80.9 and 6.2 mos.

Procedure

All *S*s were tested on three types of material (sticks, liquids, beads) for three types of quantity. Each *S* compared two amounts of each material six times. The amounts were successively arranged[1] so that twice the minimum requirement for success was a comparison of gross quantity, twice the minimum requirement for success was a comparison of intensive quantity, and twice the minimum requirement for success was a comparison of extensive quantity. *E*'s directions were always phrased with general quantity terms like "number, amount, same," which could be taken either in the gross, intensive, or extensive sense. Because of this ambiguity in language, it was assumed that the verbal directions did not provide clues to the type of quantity that had to be compared.

The school age *S*s were tested on one material at a time at weekly intervals. The pre-school *S*s were tested at hourly intervals in one morning. All *S*s took the tests in the same order; sticks, liquids, beads. The tests were structured as the games described below.

Sticks. Twenty-four wooden sticks, ¼″ square by 1¼″, were on the table. *E* "Let's pretend these sticks are candies. Your mother gives me this many candies." *E* takes six and puts them in a row at 1″ intervals. *E* "You take just as many candies as I have, take the same number of candies as me." (Comparison of intensive quantity.)

After *S* took his "candies" *E* shortened his row and asked *S* to do the same. If *S* had taken more or less than six, shortening the rows made this immediately apparent and *E* asked him to "Make them the same." When the two rows were closed and of the same length, *E* asked "Do we both have the same number of candies?" (Comparison of gross quantity.)

E then spread his "candies" apart again but left *S*'s together so that *E*'s row was much longer but still equal in number of *S*'s row. *E* "Do we both have the same number of candies?" (Comparison of extensive quantity.)

---

[1] Piaget's qualitative behavior analyses (1952) revealed what type of quantity comparison a particular arrangement required for a correct solution

The entire test was then repeated. The second time *E* arranged the sticks lengthwise and asked *S* to pretend the sticks were railroad cars and and that they were making trains. The questions were phrased accordingly, e.g., "Do we both have the same number of cars?" etc.

Liquids. Two 16 oz. and two 8 oz. drinking glasses were on the table beside a tall narrow glass and a pitcher of orange colored water. *E* "Let's pretend we are at a party and this colored water is orangeade. Your mother pours this much for you," *E* fills one 16 oz. glass ¾ full, "and this much for me," *E* fills the other 16 oz. glass ¾ full. *E* "Do we both have the same amount of orangeade to drink, do I have just as much as you?" (Comparison of gross quantity.)

Then, while *S* watched, *E* poured the "orangeade" from *E*'s glass into the two 8 oz. glasses so that the levels in the two were equal but lower than the level in the 16 oz. glass. *E* "Now my drink is in these two glasses and yours is in this glass, do I have the same amount of drink as you?" (Comparison of extensive quantity.)

Next *E* poured all the orangeade back into the pitcher and took one 16 oz. glass and the narrow glass leaving the other containers aside. *E* filled the wide and narrow glasses to the same level and said "My drink is in here," pointing to the narrow glass, "and your drink is in there," pointing to the 16 oz. glass. *E* "Do we both have the same amount to drink, do I have as much to drink as you?" (Intensive comparison.)

E then repeated the entire procedure but the second time used the 8 oz. glass in the comparison with the narrow glass.

Beads. The procedures and containers used in this test were the same as those in the test with liquids. However, large wooden beads instead of colored water were used as the material. The diameter of the beads was such that only one could fit into the narrow glass at a time making a single column like an upturned necklace. The questions were phrased "Do we have the same amount of beads, do we both have the same number of beads?" etc.

Scoring. Each successful comparison was given a score of 1 and all failures were scored zero. For each *S* there was a total possible score of 18; for each type of material and for each type of quantity there was a total possible score of six.

Statistical analysis. To test the major hypotheses of this study—about the separate and combined effects of Age, Type of Quantity, and Type of Material on children's success in comparing quantity—a 3 × 3 × 3 analysis of variance design was used. For two variables, Type of Quantity and Type of Material, between subject differences were controlled by testing all *S*s on all types of quantity and all types of material.

Table 1: *Comparison score means, standard deviations and t ratios for three age groups*

Age group	N	Mean Score	SD	t
4	18	7.78	1.51	4.15*
5	40	9.98	2.46	7.87*
6-7	22	15.41	2.67	11.39*
4	—	—	—	

** Significant beyond the .01 level.*

RESULTS[2]

Age

Piaget found that children's success in comparing quantity increased with age. In the present study the over-all *F* for the effects of Age was 56.88 and was significant beyond the .01 level. Table 1 gives the Age Group means, standard deviations, and *t* ratios. It shows that success in comparing quantities increased significantly with age in agreement with Piaget's finding.

Type of Quantity

Piaget's illustrative examples indicated that other things being equal, the type of quantity required for a correct comparison affected children's success and failure. The same held true for the *S*s of the present study. For the Type of Quantity means the over-all *F* was 480.90 and was significant beyond the .01 level. Statistics for the individual Type of Quantity means are given in Table 2. Gross quantities were the easiest to compare, intensive were intermediate, and extensive quantities were the hardest to compare. This result agreed with what Piaget reported.

---

[2] There was no reason to expect, from Piaget's findings, significant Age-Type of Material or Age-Type of Quantity-Type of Material interactions, and the *F*'s for these interactions were NS.

Table 2: *Comparison score means, standard deviations and t ratios for three types of quantity* $(N = 80)$

Type of quantity	Mean score	SD	t
Gross	5.98	0.71	3.80*
Intensive	3.51	1.83	5.12*
Extensive	1.47	2.47	5.82*
Gross	—	—	

** Significant beyond the .01 level.*

Table 3: *Comparison score means, standard deviations and t ratios for three types of material* $(N = 80)$

Type of material	Mean score	SD	t
Stick	3.84	1.40	3.26*
Liquids	2.99	1.26	4.46*
Beads	4.15	1.50	1.19
Sticks	—	—	

** Significant beyond the .01 level.*

Type of Material

In regard to the many types of material he used, Piaget concluded that success in comparing a given type of quantity of a certain type of material did not necessarily generalize to all materials. In the present study the *F* for Type of Material was 48.00 and was significant beyond the .01 level. Table 3, which gives the statistics for individual means, shows that liquids were the most difficult material to compare and produced the significant *F*. Piaget does not report this result with liquids as a specific finding but suggests comparison of a quantity of liquid may be more difficult because liquids are not broken up into perceivable units as are the sticks and beads. The present results support this assumption.

Interaction of Age with Type of Quantity

The stages Piaget reported suggested that success in comparing quantity depended jointly on age and the type of quantity comparison required by the test. In the study reported here, this joint effect was tested by the Interaction of Age with Type of Quantity. The *F* for this interaction was 35.57 and was significant beyond the .01 level. Table 4 shows that while the order of difficulty for types of quantity was the same at each age, the difference in difficulty decreased with age, and only the means at the four and five year levels were significantly different.

As indicated in Table 4, the majority of the four year group succeeded only with gross quantities, the majority of the five year group succeeded only with gross and intensive quantities, while most of the six and seven year group succeeded with all types of quantity. This agreed with Piaget's findings of age related, hierarchically ordered stages in the development of success in comparing quantity.

Interaction of Type of Quantity with Type of Material

The interaction *F* for Type of Quantity and Type of Material was 15.17 and was significant beyond the .01 level. Table 5 shows that the order of difficulty for types of quantity was the same for all materials. Table 5 also shows that the order of difficulty for types of material was the same for intensive and extensive quantities. However, the differences between the means of Type of Material were greater for intensive quantities than they were for extensive quantities. The intensive quantity of liquids, perceived in a narrow glass vs. perceived in wide glass, was unusually difficult for the *S*s of this study to compare.

Correlations for Three Types of Material

Implicit in Piaget's work on quantity was the assumption that the comparisons with different materials tapped the same conceptualizing ability. Table 6 gives the intercorrelations for the three materials used in the present study. The correlations were all high and significant. This result supported Piaget's assumption that a common conceptualizing ability underlies children's success in comparing quantity with different materials.

Comparison and W.I.S.C. Score Correlations

The Wechsler Intelligence Scale for Children was administered to the 62 school age children who took part in the study. For each age group

Table 4: *Comparison score means and F ratios for three types of quantity for three age groups*

N	Age group	Quantity compared			F
		Gross	Intensive	Extensive	
18	4	5.88	1.77	0.11	221.10*
40	5	6.0	3.20	0.75	220.76*
22	6-7	6.0	5.49	3.95	4.78

* *Significant beyond the .01 level.*

Table 5: *Comparison score means and F ratios for three types of quantity for three types of material* ($N = 80$)

Material used in comparison	Quantity compared			F
	Gross	Intensive	Extensive	
Sticks	1.99	1.38	.48	96.46*
Liquids	2.0	.66	.33	179.37*
Beads	1.99	1.48	.69	65.99*
F	.01	78.20*	13.20*	

* *Significant beyond the .01 level.*

(5 and 6–7) Z scores were computed for the quantity scores, and these were used in the correlations with the IQs given in Table 7. In the sub-test correlations obtained comparison scores were used. The correlations were generally positive, sometimes significant, and usually low. High correlations occurred with Picture Arrangement, Arithmetic and Coding which, like the comparison of quantity, involve conceptual organizing.

DISCUSSION

The results of the present study were in agreement with Piaget's finding three, age-related, hierarchically ordered stages of success in comparing quantity. Piaget assumes that success in comparing quantity earmarks developmental changes in both the form and the content of children's quantitative thinking.

With regard to content Piaget's view is that stages of success in comparing quantity earmark steps in the development of an abstract quantity concept. Children at the first stage have only a general impression of quantity which Piaget calls a *global* conception. When forced to break down this impression and compare objects, first stage children make use of single perceived relations or gross quantity. Second stage children have a differentiated, relatively precise impression of quantity which Piaget speaks of as an *intuitive* conception. In comparing objects, second stage children deal with perceived relations two by two or with intensive quantity. At the third stage children have the impression of quantity as a logical whole which Piaget speaks of as an *abstract* conception. In comparing objects, third stage children do so in unit terms or by extensive quantity.

With regard to the form of thought Piaget assumes that stages of success in comparing quantity earmarks the development of internalized

Table 6: *Comparison score correlations for three types of material* ($N = 80$)

Material	Liquids	Beads
Sticks	.90*	.64*
Liquids		.76*

* *Significant beyond the .01 level.*

Table 7: *Quantity comparison and W.I.S.C. score correlations* (*N* = 62)

Sub test correlations[a]			
Verbal		**Performance**	
Vocabulary	.31	Picture Completion	.13
Information	.47*	Picture Arrangement	.55*
Similarities	.29	Block Design	.26
Comprehension	.19	Object Assembly	.38*
Arithmetic	.35*	Coding	.42*
Digit Span	.20		

IQ Correlations[b]	
Verbal	.47*
Performance	.29
Full Scale	.43*

[a] *With obtained Comparison Scores.*
[b] *With transformed Comparison Scores.*
* *Significant beyond the .01 level.*

actions which he calls *mental operations*. It is the development of *logical multiplication* (the operation of coordinating two judgments of perceived relations) which makes possible the step from gross to intensive quantification and it is logical multiplication plus the development of *equation of differences* (the operation of coordinating judgments of the magnitude of differences between objects) which makes possible the step from intensive to extensive quantification.

The present study did not attempt to test Piaget's conception of the development of quantitative thinking in the child. It did seek to further substantiate some of those observations which Piaget reports and builds his conception upon.

SUMMARY

Eighty school and pre-school children were divided into three Age Groups (4, 5, 6–7) and tested on three Types of Material for three Types of Quantity in a systematic replication of Piaget's investigation of the development of quantitative thinking. Analysis of variance showed that success in comparing quantities varied significantly with Age, Type of Quantity, Type of Material, and two of the interactions. Correlations for Types of Material were positive, high, and significant. Correlations of comparison scores and W.I.S.C. scores were positive, generally low, and sometimes significant.

The results were in close agreement with Piaget's finding that success in comparing quantity developed in three, age related, hierarchically ordered, stages. Piaget's conception of these stages as representing steps in the development of the form and content of children's quantitative thinking was briefly presented.

REFERENCES

PIAGET, J. *Judgment and reasoning in the child*. London: Routledge & Kegan Paul, 1928.

PIAGET, J. *The child's conception of the world*. London: Routledge & Kegan Paul, 1929.

PIAGET, J. *The child's conception of physical causality*. London: Routledge & Kegan Paul, 1930.

PIAGET, J. *The child's conception of number*. London: Routledge & Kegan Paul, 1952 (first French printing, 1941).

8

The Act of Discovery

Jerome S. Bruner

Bruner, like Piaget, is a student of cognition. While Piaget is concerned with the developmental aspects of learning, Bruner has turned his attention to the educational aspects. Of special interest to Bruner are the relationships between motivation and cognitive behavior. The first two articles on motivation (1 and 2) combined with this selection offer the student a view of developmental psychology that emphasizes the organizational, goal-seeking behaviors of the growing person.

Maimonides, in his *Guide for the Perplexed* (1956), speaks of four forms of perfection that men might seek. The first and lowest form is perfection in the acquisition of worldly goods. The great philosopher dismisses such perfection on the ground that the possessions one acquires bear no meaningful relation to the possessor: "A great king may one morning find that there is no difference between him and the lowest person." A second perfection is of the body, its conformation and skills. Its failing is that it does not reflect on what is uniquely human about man: "he could [in any case] not be as strong as a mule." Moral perfection is the third, "the highest degree of excellency in man's character." Of this perfection Maimonides says: "Imagine a person being alone, and having no connection whatever with any other person; all his good moral principles are at rest, they are not required and give man no perfection whatever. These principles are only necessary and useful when man comes in contact with others." "The fourth kind of perfection is the true perfection of man; the possession of the highest intellectual faculties. . . ." In justification of his assertion, this extraordinary Spanish-Judaic philosopher urges: "Examine the first three kinds of perfection; you will find that if you possess them, they are not your property, but the property of others. . . . But the last kind of perfection is exclusively yours; no one else owns any part of it."

It is a conjecture much like that of Maimonides that leads me to examine the act of discovery in man's intellectual life. For if man's intellectual excellence is the most his own among his perfections, it is also the case that the most uniquely personal of all that he knows is that which he has discovered for himself. What difference does it make, then, that we encourage discovery in the learning of the young? Does it, as Maimonides would say, create a special and unique relation between knowledge possessed and the possessor? And what may such a unique relation do for a man—or for a child, if you will, for our concern is with the education of the young?

The immediate occasion for my concern with discovery—and I do not restrict discovery to the act of finding out something that before was unknown to mankind, but rather include all forms of obtaining knowledge for oneself by the use of one's own mind—the immediate occasion is the work of the various new curriculum projects that have grown up in America during the last six or seven years. For whether one speaks to mathematicians or physicists or historians, one encounters repeatedly an expression of faith in the powerful effects that come from permitting the student to put things together for himself, to be his own discoverer.

First, let it be clear what the act of discovery entails. It is rarely, on the frontier of knowledge or elsewhere, that new facts are "discovered" in the sense of being encountered as Newton suggested in the form of islands of truth in an uncharted sea of ignorance. Or if they appear to be discovered in this way, it is almost always thanks to some happy hypotheses about where to navigate. Discovery, like surprise, favors the well prepared mind. In playing bridge, one is surprised by a hand with no honors in it at all and also by hands that are all in one suit. Yet all hands in bridge are equiprobable: one must know to be surprised. So too in discovery. The history of science is studded with examples of men "finding out" something and not knowing it. I shall operate on the assumption that discovery, whether by a schoolboy going it on his own or by a scientist cultivating the growing edge of his field, is in its essence a matter of rearranging or transforming evidence in such a way that one is enabled to go beyond the evidence so reassembled to additional new insights. It may well be that an additional fact or shred of evidence makes this larger transformation of evidence possible. But it is often not even dependent on new information.

From *Harvard Education Review*, 1961, *31*: 21–32.

It goes without saying that, left to himself, the child will go about discovering things for himself within limits. It also goes without saying that there are certain forms of child rearing, certain home atmospheres that lead some children to be their own discoverers more than other children. These are both topics of great interest, but I shall not be discussing them. Rather, I should like to confine myself to the consideration of discovery and "finding-out-for-oneself" within an educational setting—specifically the school. Our aim as teachers is to give our student as firm a grasp of a subject as we can, and to make him as autonomous and self-propelled a thinker as we can—one who will go along on his own after formal schooling has ended. I shall return in the end to the question of the kind of classroom and the style of teaching that encourages an attitude of wanting to discover. For purposes of orienting the discussion, however, I would like to make an overly simplified distinction between teaching that takes place in the *expository mode* and teaching that utilizes the *hypothetical mode*. In the former, the decisions concerning the mode and pace and style of exposition are principally determined by the teacher as expositor; the student is the listener. If I can put the matter in terms of structural linguistics, the speaker has a quite different set of decisions to make than the listener: the former has a wide choice of alternatives for structuring, he is anticipating paragraph content while the listener is still intent on the words, he is manipulating the content of the material by various transformations, while the listener is quite unaware of these internal manipulations. In the hypothetical mode, the teacher and the student are in a more cooperative position with respect to what in linguistics would be called "speaker's decisions." The student is not a bench-bound listener, but is taking a part in the formulation and at times may play the principle role in it. He will be aware of alternatives and may even have an "as if" attitude toward these and, as he receives information he may evaluate it as it comes. One cannot describe the process in either mode with great precision as to detail, but I think the foregoing may serve to illustrate what is meant.

Consider now what benefit might be derived from the experience of learning through discoveries that one makes for oneself. I should like to discuss these under four headings: (1) The increase in intellectual potency, (2) the shift from extrinsic to intrinsic rewards, (3) learning the heuristics of discovering, and (4) the aid to memory processing.

INTELLECTUAL POTENCY

If you will permit me, I would like to consider the difference between subjects in a highly constrained psychological experiment involving a two-choice apparatus. In order to win chips, they must depress a key either on the right or the left side of the machine. A pattern of payoff is designed such that, say, they will be paid off on the right side 70 per cent of the time, on the left 30 per cent, although this detail is not important. What is important is that the payoff sequence is arranged at random, and there is no pattern. I should like to contrast the behavior of subjects who think that there *is* some pattern to be found in the sequence—who think that regularities are discoverable—in contrast to subjects who think that things are happening quite by *chance*. The former group adopts what is called an "event-matching" strategy in which the number of responses given to each side is roughly equal to the proportion of times it pays off: in the present case R70 : L30. The group that believes there is no pattern very soon reverts to a much more primitive strategy wherein *all* responses are allocated to the side that has the greater payoff. A little arithmetic will show you that the lazy all-and-none strategy pays off more if indeed the environment is random: namely, they win seventy per cent of the time. The event-matching subjects win about 70% on the 70% payoff side (or 49% of the time there) and 30% of the time on the side that pays off 30% of the time (another 9% for a total take-home wage of 58% in return for their labors of decision). But the world is not always or not even frequently random, and if one analyzes carefully what the event-matchers are doing, it turns out that they are trying out hypotheses one after the other, all of them containing a term such that they distribute bets on the two sides with a frequency to match the actual occurrence of events. If it should turn out that there is a pattern to be discovered, their payoff would become 100%. The other group would go on at the middling rate of 70%.

What has this to do with the subject at hand? For the person to search out and find regularities and relationships in his environment, he must be armed with an expectancy that there will be something to find and, once aroused by expectancy, he must devise ways of searching and finding. One of the chief enemies of such expectancy is the assumption that there is nothing one can find in the environment by way of regularity or relationship. In the experiment just cited, subjects often fall into a habitual attitude that there is either nothing to be found or that they can find a pattern by looking. There is an important sequel in behavior to the two attitudes, and to this I should like to turn now.

We have been conducting a series of experimental studies on a group of some seventy school children over the last four years. The studies have led us to distinguish an interesting dimension of cognitive activity that can be described as ranging from *episodic empiricism* at one end to *cumulative constructionism* at the other. The two attitudes in the choice experiments just cited are illustrative of the extremes of the dimension. I might mention some other illustrations. One of the experiments employs the game of Twenty Questions. A child—in this case he is between 10 and 12—is told that a car has gone off the road and hit a tree. He is to ask questions that can be answered by "yes" or "no" to discover the cause of the accident. After completing the problem, the same task is given him again, though he is told that the accident had a different cause this time. In all, the procedure is repeated four times. Children enjoy playing the game. They also differ quite markedly in the approach or strategy they bring to the task. There are various elements in the strategies employed. In the first place, one may distinguish clearly between two types of questions asked: the one is designed for locating constraints in the problem, constraints that will eventually give shape to an hypothesis; the other is the hypothesis as question. It is the difference between, "Was there anything wrong with the driver?" and "Was the driver rushing to the doctor's office for an appointment and the car got out of control?" There are children who precede hypotheses with efforts to locate constraint and there are those who, to use our local slang, are "pot-shotters," who string out hypotheses noncumulatively one after the other. A second element of strategy is its connectivity of information gathering: the extent to which questions asked utilize or ignore or violate information previously obtained. The questions asked by children tend to be organized in cycles, each cycle of questions usually being given over to the pursuit of some particular notion. Both within cycles and between cycles one can discern a marked difference on the connectivity of the child's performance. Needless to say, children who employ constraint location as a technique preliminary to the formulation of hypotheses tend to be far more connected in their harvesting of information. Persistence is another feature of strategy, a characteristic compounded of what appear to be two components: a sheer doggedness component, and a persistence that stems from the sequential organization that a child brings to the task. Doggedness is probably just animal spirits or the need for achievement—what has come to be called *n-ach*. Organized persistence is a maneuver for protecting our fragile cognitive apparatus from overload. The child who has flooded himself with disorganized information from unconnected hypotheses will become discouraged and confused sooner than the child who has shown a certain cunning in his strategy of getting information—a cunning whose principal component is the recognition that the value of information is not simply in getting it but in being able to carry it. The persistence of the organized child stems from his knowledge of how to organize questions in cycles, how to summarize things to himself, and the like.

Episodic empiricism is illustrated by information gathering that is unbound by prior constraints, that lacks connectivity, and that is deficient in organizational persistence. The opposite extreme is illustrated by an approach that is characterized by constraint sensitivity, by connective maneuvers, and by organized persistence. Brute persistence seems to be one of those gifts from the gods that make people more exaggeratedly what they are.[1]

Before returning to the issue of discovery and its role in the development of thinking, let me say a word more about the ways in which information may get transformed when the problem solver has actively processed it. There is first of all a pragmatic question: what does it take to get information processed into a form best designed to fit some future use? Take an experiment by Zajonc[2] as a case in point. He gives groups of subjects information of a controlled kind, some groups being told that their task is to transmit the information to others, others that it is merely to be kept in mind. In general, he finds more differentiation and organization of the information received with the intention of being transmitted than there is for information received passively. An active set leads to a transformation related to a task to be performed. The risk, to be sure, is in possible overspecialization of information processing that may lead to such a high degree of specific organization that information is lost for general use.

---

[1] I should also remark in passing that the two extremes also characterize concept attainment strategies as reported in *A Study of Thinking* by J. S. Bruner *et al.* (1956). Successive scanning illustrates well what is meant here by episodic empiricism; conservative focussing is an example of cumulative constructionism.

[2] R. B. Zajonc (Personal communication, 1957).

I would urge now in the spirit of an hypothesis that emphasis upon discovery in learning has precisely the effect upon the learner of leading him to be a constructionist, to organize what he is encountering in a manner not only designed to discover regularity and relatedness, but also to avoid the kind of information drift that fails to keep account of the uses to which information might have to be put. It is, if you will, a necessary condition for learning the variety of techniques of problem solving, of transforming information for better use, indeed for learning how to go about the very task of learning. Practice in discovering for oneself teaches one to acquire information in a way that makes that information more readily viable in problem solving. So goes the hypothesis. It is still in need of testing. But it is an hypothesis of such important human implications that we cannot afford not to test it—and testing will have to be in the schools.

INTRINSIC AND EXTRINSIC MOTIVES

Much of the problem in leading a child to effective cognitive activity is to free him from the immediate control of environmental rewards and punishments. That is to say, learning that starts in response to the rewards of parental or teacher approval or the avoidance of failure can too readily develop a pattern in which the child is seeking cues as to how to conform to what is expected of him. We know from studies of children who tend to be early over-achievers in school that they are likely to be seekers after the "right way to do it" and that their capacity for transforming their learning into viable thought structures tends to be lower than children merely achieving at levels predicted by intelligence tests. Our tests on such children show them to be lower in analytic ability than those who are not conspicuous in overachievement (Bruner & Caron, in press). As we shall see later, they develop rote abilities and depend upon being able to "give back" what is expected rather than to make it into something that relates to the rest of their cognitive life. As Maimonides would say, their learning is not their own.

The hypothesis that I would propose here is that to the degree that one is able to approach learning as a task of discovering something rather than "learning about" it, to that degree will there be a tendency for the child to carry out his learning activities with the autonomy of self-reward or, more properly by reward that is discovery itself.

To those of you familiar with the battles of the last half-century in the field of motivation, the above hypothesis will be recognized as controversial. For the classic view of motivation in learning has been, until very recently, couched in terms of a theory of drives and reinforcement: that learning occurred by virtue of the fact that a response produced by a stimulus was followed by the reduction in a primary drive state. The doctrine is greatly extended by the idea of secondary reinforcement: any state associated even remotely with the reduction of a primary drive could also have the effect of producing learning. There has recently appeared a most searching and important criticism of this position, written by Professor Robert White (1956),* reviewing the evidence of recently published animal studies, of work in the field of psychoanalysis, and of research on the development of cognitive processes in children. Professor White comes to the conclusion, quite rightly I think, that the drive-reduction model of learning runs counter to too many important phenomena of learning and development to be either regarded as general in its applicability or even correct in its general approach. Let me summarize some of his principal conclusions and explore their applicability to the hypothesis stated above.

> I now propose that we gather the various kinds of behavior just mentioned, all of which have to do with effective interaction with the environment, under the general heading of competence. According to Webster, competence means fitness or ability, and the suggested synonyms include capability, capacity, efficiency, proficiency, and skill. It is therefore a suitable word to describe such things as grasping and exploring, crawling and walking, attention and perception, language and thinking, manipulating and changing the surroundings, all of which promote an effective—a competent—interaction with the environment. It is true of course, that maturation plays a part in all these developments, but this part is heavily overshadowed by learning in all the more complex accomplishments like speech or skilled manipulation. I shall argue that it is necessary to make competence a motivational concept; there is *competence motivation* as well as competence in its more familiar sense of achieved

---

* This article appears in this volume—*Editor's note*

> capacity. The behavior that leads to the building up of effective grasping, handling, and letting go of objects, to take one example, is not random behavior that is produced by an overflow of energy. It is directed, selective, and persistent, and it continues not because it serves primary drives, which indeed it cannot serve until it is almost perfected, but because it satisfies an intrinsic need to deal with the environment (White, 1956, pp. 317–318).

I am suggesting that there are forms of activity that serve to enlist and develop the competence motive, that serve to make it the driving force behind behavior. I should like to add to White's general premise that the *exercise* of competence motives has the effect of strengthening the degree to which they gain control over behavior and thereby reduce the effects of extrinsic rewards or drive gratification.

The brilliant Russian psychologist Vigotsky (1934) characterizes the growth of thought processes as starting with a dialogue of speech and gesture between child and parent; autonomous thinking begins at the stage when the child is first able to internalize these conversations and "run them off" himself. This is a typical sequence in the development of competence. So too in instruction. The narrative of teaching is of the order of the conversation. The next move in the development of competence is the internalization of the narrative and its "rules of generation" so that the child is now capable of running off the narrative on his own. The hypothetical mode in teaching by encouraging the child to participate in "speaker's decisions" speeds this process along. Once internalization has occurred, the child is in a vastly improved position from several obvious points of view—notably that he is able to go beyond the information he has been given to generate additional ideas that can either be checked immediately from experience or can, at least, be used as a basis for formulating reasonable hypotheses. But over and beyond that, the child is now in a position to experience success and failure not as reward and punishment, but as information. For when the task is his own rather than a matter of matching environmental demands, he becomes his own paymaster in a certain measure. Seeking to gain control over his environment, he can now treat success as indicating that he is on the right track, failure as indicating he is on the wrong one.

In the end, this development has the effect of freeing learning from immediate stimulus control. When learning in the short run leads only to pellets of this or that rather than to mastery in the long run, then behavior can be readily "shaped" by extrinsic rewards. When behavior becomes more long-range and competence-oriented, it comes under the control of more complex cognitive structures, plans and the like, and operates more from the inside out. It is interesting that even Pavlov, whose early account of the learning process was based entirely on a notion of stimulus control of behavior through the conditioning mechanism in which, through contiguity a new conditioned stimulus was substituted for an old unconditioned stimulus by the mechanism of stimulus substitution, that even Pavlov recognized his account as insufficient to deal with higher forms of learning. To supplement the account, he introduced the idea of the "second signalling system," with central importance placed on symbolic systems such as language in mediating and giving shape to mental life. Or as Luria (1959) has put it, "the first signal system [is] concerned with directly perceived stimuli, the second with systems of verbal elaboration." Luria, commenting on the importance of the transition from first to second signal system, says: "It would be mistaken to suppose that verbal intercourse with adults merely changes the contents of the child's conscious activity without changing its form. . . . The word has a basic function not only because it indicates a corresponding object in the external world, but also because it abstracts, isolates the necessary signal, generalizes perceived signals and relates them to certain categories; it is this systematization of direct experience that makes the role of the word in the formation of mental processes so exceptionally important" (Luria, 1959, p. 12).[3]

It is interesting that the final rejection of the universality of the doctrine of reinforcement in direct conditioning came from some of Pavlov's own students. Ivanov-Smolensky (1951) and Krasnogorsky (1954) published papers showing the manner in which symbolized linguistic messages could take over the place of the unconditioned stimulus and of the unconditioned response (gratification of hunger) in children. In all instances, they speak of these as *replacements* of lower, first-system mental or neural processes by higher order or second-system controls. A strange irony, then, that Russian psychology that gave us the notion of the conditioned response and the assumption that higher order activities are built up out of colligations or structurings of such primitive units, rejected this notion while much of American learning psychology has stayed until quite recently within the early Pavlovian fold (see, for example, a recent article by Spence (1959) in the *Harvard Educational Review* or Skinner's treatment of language (1957) and the attacks that have been made upon it by linguists such as Chomsky (1957) who have become concerned with the relation of language and cognitive activity). What is the more interesting is that Russian pedagogical theory has become deeply influenced by this new trend and is now placing much stress upon the importance of building up a more active symbolical approach to problem solving among children.

---

[3] For an elaboration of the view expressed by Luria, the reader is referred to the translation of L. S. Vigotsky's 1934 book published by John Wiley and Sons and the Technology Press, 1961.

To sum up the matter of the control of learning, then, I am proposing that the degree to which competence or mastery motives come to control behavior, to that degree the role of reinforcement or "extrinsic pleasure" wanes in shaping behavior. The child comes to manipulate his environment more actively and achieves his gratification from coping with problems. Symbolic modes of representing and transforming the environment arise and the importance of stimulus-response-reward sequences declines. To use the metaphor that David Riesman developed in a quite different context, mental life moves from a state of outer-directedness in which the fortuity of stimuli and reinforcement are crucial to a state of inner-directedness in which the growth and maintenance of mastery become central and dominant.

LEARNING THE HEURISTICS OF DISCOVERY

Lincoln Steffens (1931), reflecting in his *Autobiography* on his undergraduate education at Berkeley, comments that his schooling was overly specialized on learning about the known and that too little attention was given to the task of finding out about what was not known. But how does one train a student in the techniques of discovery? Again I would like to offer some hypotheses. There are many ways of coming to the arts of inquiry. One of them is by careful study of its formalization in logic, statistics, mathematics, and the like. If a person is going to pursue inquiry as a way of life, particularly in the sciences, certainly such study is essential. Yet, whoever has taught kindergarten and the early primary grades or has had graduate students working with him on their theses—I choose the two extremes for they are both periods of intense inquiry—knows that an understanding of the formal aspect of inquiry is not sufficient. There appear to be, rather, a series of activities and attitudes, some directly related to a particular subject and some of them fairly generalized, that go with inquiry and research. These have to do with the *process* of trying to find out something and while they provide no guarantee that the *product* will be any *great* discovery, their absence is likely to lead to awkwardness or aridity or confusion. How difficult it is to describe these matters—the heuristics of inquiry. There is one set of attitudes or ways of doing that has to do with sensing the relevance of variables—how to avoid getting stuck with edge effects and getting instead to the big sources of variance. Partly this gift comes from intuitive familiarity with a range of phenomena, sheer "knowing the stuff." But it also comes out of a sense of what things among an ensemble of things "smell right" in the sense of being of the right order of magnitude or scope or severity.

The English philosopher Weldon describes problem solving in an interesting and picturesque way. He distinguishes between difficulties, puzzles, and problems. We solve a problem or make a discovery when we impose a puzzle form on to a difficulty that converts it into a problem that can be solved in such a way that it gets us where we want to be. That is to say, we recast the difficulty into a form that we know how to work with, then work it. Much of what we speak of as discovery consists of knowing how to impose what kind of form on various kinds of difficulties. A small part but a crucial part of discovery of the highest order is to invent and develop models or "puzzle forms" that can be imposed on difficulties with good effect. It is in this area that the truly powerful mind shines. But it is interesting to what degree perfectly ordinary people can, given the benefit of instruction, construct quite interesting and what, a century ago, would have been considered greatly original models.

Now to the hypothesis. It is my hunch that it is only through the exercise of problem solving and the effort of discovery that one learns the working heuristic of discovery, and the more one has practice, the more likely is one to generalize what one has learned into a style of problem solving or inquiry that serves for any kind of task one may encounter—or almost any kind of task. I think the matter is self-evident, but what is unclear is what kinds of training and teaching produce the best effects. How do we teach a child to, say, cut his losses but at the same time be persistent in trying out an idea; to risk forming an early hunch without at the same time formulating one *so* early and with so little evidence as to be stuck with it waiting for appropriate evidence to materialize; to pose good testable guesses that are neither too brittle nor too sinuously incorrigible; etc., etc. Practice in inquiry, in trying to figure out things for oneself is indeed what is needed, but in what form? Of only one thing I am convinced. I have never seen anybody improve in the art and technique of inquiry by any means other than engaging in inquiry.

CONSERVATION OF MEMORY

I should like to take what some psychologists might consider a rather drastic view of the memory process. It is a view that in large measure derives from the work of my colleague, Professor George Miller (1956). Its first premise is that the principal problem of human memory is not storage, but retrieval. In spite of the biological unlikeliness of it, we seem to be able to store a huge quantity of information—perhaps not a full tape recording, though at times it seems we even do that, but a great sufficiency of impressions. We may infer this from the fact that recognition (i.e., recall with the aid of maximum prompts) is so extraordinarily good in human beings—particularly in comparison with spontaneous recall where, so to speak, we must get out stored information without external aids or prompts. The key to retrieval is organization or, in even simpler terms, knowing where to find information and how to get there.

Let me illustrate the point with a simple experiment. We present pairs of words to twelve-year-old children. One group is simply told to remember the pairs, that they will be asked to repeat them later. Another is told to remember them by producing a word or idea that will tie the pair together in a way that will make sense to them. A third group is given the mediators used by the second group when presented with the pairs to aid them in tying the pairs into working units. The word pairs include such juxtapositions as "chair-forest," "sidewalk-square," and the like. One can distinguish three styles of mediators and children can be scaled in terms of their relative preference for each: *generic mediation* in which a pair is tied together by a superordinate idea: "chair and forest are both made of wood"; *thematic mediation* in which the two terms are imbedded in a theme or little story: "the lost child sat on a chair in the middle of the forest"; and *part-whole mediation* where "chairs are made from trees in the forest" is typical. Now, the chief result, as you would all predict, is that children who provide their own mediators do best—indeed, one time through a set of thirty pairs, they recover up to 95% of the second words when presented with the first ones of the pairs, whereas the uninstructed children reach a maximum of less than 50% recovered. Interestingly enough, children do best in recovering materials tied together by the form of mediator they most often use.

One can cite a myriad of findings to indicate that any organization of information that reduces the aggregate complexity of material by imbedding it into a cognitive structure a person has constructed will make that material more accessible for retrieval. In short, we may say that the process of memory, looked at from the retrieval side, is also a process of problem solving: how can material be "placed" in memory so that it can be got on demand?

We can take as a point of departure the example of the children who developed their own technique for relating the members of each word pair. You will recall that they did better than the children who were given by exposition the mediators they had developed. Let me suggest that in general, material that is organized in terms of a person's own interests and cognitive structures is material that has the best chance of being accessible in memory. That is to say, it is more likely to be placed along routes that are connected to one's own ways of intellectual travel.

In sum, the very attitudes and activities that characterize "figuring out" or "discovering" things for oneself also seems to have the effect of making material more readily accessible in memory.

REFERENCES

BRUNER, J. S., & CARON, A. J. Cognition, anxiety, and achievement in the preadolescent. *J. educ. Psychol.*, in press.

BRUNER, J. S., *et al. A study of thinking.* New York: Wiley, 1956.

CHOMSKY, N. *Syntactic structure.* The Hague, The Netherlands: Mouton & Co., 1957.

IVANOV-SMOLENSKY, A. G. Concerning the study of the joint activity of the first and second signal systems. *J. higher nerv. Activity*, 1951, *1*, 1.

KRASNOGORSKY, N. D. *Studies of higher nervous activity in animals and in man*. Vol. I. Moscow, 1954.

LURIA, A. L. The directive function of speech in development and dissolution. *Word*, 1959, *15*, 341–464.

MAIMONIDES. *Guide for the perplexed*. New York: Dover Publications, 1956.

MILLER, G. A. The magical number seven, plus or minus two. *Psychol. Rev.*, 1956, *63*, 81–97.

SKINNER, B. F. *Verbal behavior*. New York: Appleton-Century-Crofts, 1957.

SPENCE, K. W. The relation of learning theory to the technique of education. *Harvard educ. Rev.*, 1959, *29*, 84–95.

STEFFENS, L. *Autobiography of Lincoln Steffens*. New York: Harcourt, Brace, 1931.

VIGOTSKY, L. S. *Thinking and speech*. Moscow, 1934.

WHITE, R. W. Motivation reconsidered: The concept of competence. *Psychol. Rev.*, 1959, *66*, 297–333.

9

Sex-Role and Parental Identification

David B. Lynn

The concept of identification has been a key one in both psychoanalytic and general approaches to personality development. In this article Lynn reviews the literature on the concept. Instead of approaching it from an eclectic view, however, he sets out to test particular hypotheses about sex differences in learning to identify with parents and other hypotheses about sex differences in the approach to learning in general.

It is doubtful that psychological theories have fully posed much less resolved the question of the extent of sex differences in personality development. In this connection Sarason, *et al.* comment: "No one to our knowledge has denied they [such sex differences] are pervasive, and yet the problem of degree of pervasiveness has not been critically examined despite its implications for theory, methodology, and the direction of future research" (1960, p. 260).

A perusal of the journals shows that many studies, which include both male and female *S*s in the sample, do not make provisions for sex differences in the hypotheses. Where sex differences are found they are, consequently, rationalized *post facto*. Moreover, often no statistical analysis of sex differences is performed, despite their importance in psychological processes.

This paper presents a theoretical formulation which postulates basic sex differences in the *nature* of sex-role and parental identification, as well as basic differences in the *process of achieving* such identification. The developmental processes described are considered neither inevitable nor universal. If they are appropriate to the U.S. culture today, they may, nevertheless, be inappropriate for many other cultures and for a significantly altered U.S. culture of the future. This formulation refers to the "typical" pattern, although recognizing that a "typical" pattern, if not a myth, is at least an exception. Research findings considered relevant to this formulation are reviewed.

Before developing this formulation, let us briefly define identification as it is used here. *Sex-role identification* refers to the internalization of the role considered appropriate to a given sex and to the unconscious reactions characteristic of that role. *Parental identification* refers to the internalization of personality characteristics of one's own parent and to unconscious reactions similar to that parent. Thus, theoretically, an individual might be well identified with the appropriate sex-role generally and yet poorly identified with his same-sex parent specifically. This differentiation also allows for the converse circumstances wherein a person is well identified with his same-sex parent specifically and yet poorly identified with the appropriate sex-role generally. In such an instance the parent with whom the individual is well identified is himself poorly identified with the appropriate sex-role. An example might be a girl who is well identified with her own mother, but the mother is identified with the masculine rather than the feminine role. Such a girl, therefore, through her identification with her mother, is poorly identified with the feminine role.

In a previous paper (1959) the author differentiated the concept of *sex-role identification* from

---

From *Child Development*. 1962, *33*: 555–564. Reprinted with permission of the Society for Research in Child Development.

sex-role preference and *sex-role adoption.* The present formulation is a departure from that previous paper and also shares various features in common with others (Brown, 1956; Douvan, 1957; Hartley, 1959; Sears, Maccoby, & Levin, 1957).

This formulation uses a hypothesis from the previous paper as a postulate from which to deduce a number of new hypotheses. Hopefully, this formulation will offer a unified theoretical framework consistent with a number of varied findings concerning sex differences.

The aspects of the previous formulation pertinent to the present one are summarized as follows:

Both male and female infants were hypothesized to learn to identify with the mother. Boys, but not girls, must shift from this initial identification with the mother to masculine identification. The girl has the same-sex parental model for identification (the mother) with her more than the boy has the same-sex model (the father) with him. Much incidental learning takes place from the girl's contact with her mother which she can apply directly in her life.

However, despite the shortage of male models, a somewhat stereotyped and conventional masculine role is nonetheless spelled out for the boys, e.g., by his mother and women teachers in the absence of his father and male teachers. In this connection a study by Sherriffs and Jarrett (1953) indicated that men and women share the same stereotypes about the two sexes. Through the reinforcement of the culture's highly developed system of rewards for indications of masculinity and punishment for signs of femininity, the boy's early learned identification with the mother eventually weakens and becomes more or less replaced by the later learned identification with a culturally defined, somewhat stereotyped masculine role. *"Consequently, males tend to identify with a cultural stereotype of the masculine role, whereas females tend to identify with aspects of their own mothers' role specifically"* (Lynn, 1959, p. 130). This hypothesis was generally supported by the research findings reviewed (Gray & Klaus, 1956, Lazowick, 1955).

This hypothesis is not meant to minimize the role of the father in the development of males. Studies of father-absence suggest that the presence of the father in the home is of great importance for boys (Bach, 1946; Lynn & Sawrey, 1959; Sears, Pintler, & Sears, 1946). It is beyond the scope of this paper to elaborate on the role of the father, but it is our position that it has a very different place in the development of the boy's masculine-role identification than does the mother in the girl's mother identification. The father, as a model for the boy, may be thought of as analogous to a map showing the major outline but lacking most details, whereas the mother, as a model for the girl, might be thought of as a detailed map. The father, of course, serves many other functions besides that of model for the boy's masculine-role identification. He may, for example, reinforce the boy's masculine strivings and stimulate his drive to achieve masculine-role identification. Because fathers typically do spend so much time away from home and, even when home, usually do not participate in as many intimate activities with the child as does the mother (e.g., preparation for bed), it is probably true that the time spent with the father takes on much importance in the boy's identification development.

Although recognizing the contribution of the father in the identification of males and the general cultural influences in the identification of females, it nevertheless seems meaningful, for simplicity in developing this formulation, to refer to *masculine-role identification* in males as distinguished from *mother identification* in females.

It is postulated that the task of achieving these separate kinds of identification for each sex requires separate methods of learning. These separate identification tasks seem to parallel the two kinds of learning tasks differentiated by Woodworth and Schlosberg: the *problem* and the *lesson.* "With a problem to master the learner must explore the situation and find the goal before his task is fully presented. In the case of a lesson, the problem-solving phase is omitted or at least minimized, as we see when the human subject is instructed to memorize this poem or that list of nonsense syllables, to examine these pictures with a view to recognizing them later. . . ." (1954, p. 529). The task of achieving mother identification for the female is considered roughly parallel to the learning *lesson,* and the task of achieving masculine-role identification for the male is considered roughly parallel to the learning *problem.*

It is assumed that finding the goal does not constitute a major problem for the girl in learning her mother identification lesson. Since the girl, unlike the boy, need not shift from the initial mother identification and since she typically has the mother with her a relatively large proportion of the time, it is postulated that the question of the object of identification (the mother) for the girl seldom arises. She learns the mother identification lesson in the context of an intimate personal relationship with the mother, partly by imitation, which as used here includes covert practice of the actions characteristic of the mother (Maccoby, 1959). She also learns the mother identification lesson through the mother's selective reinforcement of mother-similar tendencies in the girl. Hartup (1962) did a relevant study concerning parental imitation in children aged 3 to 5 in which he correlated sex-role preference in the Brown It Scale (Brown, 1956) with the degree to which the *S*'s doll play showed the child doll imitating the same-sex parental doll. The results suggested to Hartup that girls become feminine partly as a result of a tendency to imitate their mothers more than their fathers and that the acquisition of masculinity by boys appears to be independent of the tendency to imitate the father more than the mother.

Similarly, abstracting principles defining mother identification is not considered a concern for the girls. Any bit of behavior on the mother's part may be of potential importance in learning the mother identification lesson, and therefore the girl need not abstract principles defining the feminine role. It is not principles defining the feminine role that the girl need learn, but rather an identification with her specific mother.

It is assumed, on the other hand, that finding the goal *does* constitute a major problem for the boy in solving the masculine-role identification problem. There is evidence to indicate that between two-thirds and three-fourths of children by the age of 3 are able to make the basic distinction between sexes (Gesell, *et al.*, 1940; Gesell, Ilg, *et al.*, 1943; Seward, 1946). When the boy begins to be aware that he does not belong in the same sex-category as the mother, he must then find the proper sex-role identification goal. Hartley says, of the identification problem that faces the boy, ". . . the desired behavior is rarely defined positively as something the child *should* do, but rather negatively as something he should *not* do or be—anything, that is, that the parent or other people regard as 'sissy.' Thus, very early in life the boy must either stumble on the right path or bear repeated punishment without warning when he accidentally enters into the wrong ones" (1959, p. 458). From these largely negative admonishings, often made by women and often without the benefit of the presence of a male model during most of his waking hours, the boy must learn to set the masculine role as his goal. He must also restructure the admonishings, often negatively made and given in many contexts, in order to abstract the principles defining the masculine role.

One of the basic steps in this formulation can now be taken. It is assumed that, in learning the appropriate identification, each sex is thereby acquiring separate methods of learning which are subsequently applied to learning tasks generally. The little girl acquires a learning method which primarily involves: (a) a personal relationship and (b) imitation rather than restructuring the field and abstracting principles. On the other hand, the little boy acquires a different learning method which primarily involves: (a) defining the goal; (b) restructuring the field; and (c) abstracting principles.

HYPOTHESES

The following hypotheses are considered to follow from the above formulation:

1. It is in the context of a close personal relationship with the mother that the little girl learns the mother identification lesson. She is reinforced by appropriate rewards for signs that she is learning this lesson. Since the little girl is rewarded in the context of the personal relationship with her mother, maintaining the rewarding relationship with her mother should acquire strong secondary-drive characteristics. By generalization, the need for affiliation in other situations should also have strong secondary-drive characteristics for the girl.

The boy, relative to the girl, has little opportunity to receive rewards for modeling an adult male in a close personal relationship. He receives his rewards for learning the appropriate principles of masculine-role identification as they are abstracted from many contexts. Therefore, the need for affiliation in general should not acquire as much strength as a secondary drive for males as for females. *Consequently, females will tend to demonstrate greater need for affiliation than males.*

2. In learning to identify with the mother, any bit of behavior on the mother's part might be of potential importance in the girl's perception of her. The mother identification lesson does not require that the girl deviate from the given, but rather that she learn the lesson as presented.

For the boys, solving the problem of masculine-role identification must be accomplished without adequate exposure to adult male models. It must be solved by using the admonishings, such as "don't be a sissy," which, occurring in many contexts, serve as guides in defining the masculine role. To solve the masculine-role identification problem the boy must restructure the field. Therefore, the masculine learning method *does* include restructuring the field as a learning principle. *Consequently, females tend to be more dependent than males on the external context of a perceptual situation and hesitate to deviate from the given.*

3. In the process of solving the masculine-role identification problem, the male acquires a method of learning which should be applicable in solving other problems. On the other hand, the feminine learning method, emerging from the process of learning the mother identification lesson, is not well geared to problem-solving. *Consequently, males tend to surpass females in problem-solving skills.*

4. The masculine learning method is postulated to include abstracting principles, whereas the feminine one is not. The tendency to abstract principles should generalize to other problems in addition to the problem of achieving masculine-role identification. It should, for example, generalize to the acquisition of moral standards. If one is very responsive to the moral standards of others, it is relatively unnecessary to internalize standards. If one, on the other hand, tends to learn moral standards by abstracting moral principles rather than being highly responsive to the standards of others, then one *does* need to internalize one's standards. If one is to stick by one's principles, they had better be internalized. It is postulated that males more than females will tend to learn moral standards by abstracting moral principles. *Consequently, males tend to be more concerned with internalized moral standards than females.*

5. Conversely, the feminine learning method indicates that one learns by imitation through a relationship whereas the masculine learning method does not. The little girl, it was assumed, tends to learn the lesson as given, without restructuring. Such a learning method should generalize to the acquisition of standards. *Consequently, females tend to be more receptive to the standards of others than males.*

RELEVANT FINDINGS

Let us now see how consistently these hypotheses correspond to previous findings and whether this formulation helps clarify and unify the data.

Hypothesis 1, predicting that females will demonstrate greater need for affiliation than males, is supported by a study by Edwards (1959) which showed that women have significantly higher means than men on affiliation on the Edwards Personal Preference Schedule (EPPS).

McClelland, Atkinson, Clark, and Lowell (1953) found that college women did not show an increase in achievement motive scores as a result of the arousal instruction, based on reference to leadership and intelligence, effective for male college students. Women did obtain higher scores, however, when the dimension of "achievement" was social acceptability.

Lansky, Crandall, Kagan, and Baker (1961), in a study of children aged 13 to 18, used the French Insight Test (French, 1958) to measure affiliation. The French test, as used in this study, consisted of 20 items describing a characteristic behavior of a boy (girl), e.g., "Tom never joins clubs or social groups." For each item the *S* answers these three questions: (a) what is the boy (girl) like? (b) what does he (she) want to have or do? and (c) what are the results of his (her) behavior apt to be? Girls were significantly higher than boys on preoccupation with *affiliation*, which was scored when the goal is to be liked or accepted by others or to be part of a group.

When Harris (1959) repeated Symonds' 1935 studies (1936a, 1936b) of having adolescents rank interests, he found that girls persist in their greater interest in social relations than boys. When Winkler (1949) analyzed the replies of children aged 7 to 16, he found that girls seemed more interested than boys in social relationships, especially face-to-face contacts. The girl's early preoccupation with affiliation was noted by Goodenough (1957) who found that nursery school girls drew more pictures of persons and mentioned persons more often than boys.

Thus, these data are consistent with hypothesis 1 that females will demonstrate greater need for affiliation than males.

Evidence concerning hypothesis 2, predicting that females tend to be more dependent than males on the external context of a perceptual situation and will hesitate to deviate from the given, is furnished by Witkin, Lewis, Hertzman, Machover, Meissner, and Wapner (1954). They found that female *S*s were more readily influenced by misleading cues than were male *S*s and thus were higher in "perceptual-field dependence."

Additional evidence concerning this hypothesis is found in a study by Wallach and Caron (1959) with sixth-grade school children. These children were given a concept attainment session in which to establish criteria concerning geometric forms with certain characteristics. A test session followed in which the *S*s judged whether figures of varying deviation from the standard were similar to it. It was found that girls tolerated less deviation than males by every index, thus agreeing with the hypothesis that females more than males hesitate to "move away" from the given. Both studies are in agreement with hypothesis 2.

Studies reported by Sweeney (1953) are relevant to hypothesis 3 that males generally surpass females in problem-solving skills. Most of the studies reported by Sweeney support this hypothesis. Moreover, he reported experiments of his own which demonstrate that men solve certain classes of problems with greater facility than do women, even when differences in intellectual aptitude, special knowledge or training, and special abilities are controlled. Sweeney obtained scores on the College Board Scholastic Aptitude test for 130 men and 139 women to whom McNemar (1954) had given four tests of logical reasoning: False Premises, Essential Operations, Syllogisms, and Problem Solving. Significant differences favoring the men were found on all four of these tests for 100 pairs who had been matched in verbal aptitude scores. For 90 pairs matched in mathematical aptitude and 69 pairs matched both in verbal and mathematical aptitude, a difference was obtained only for Problem Solving, a test which essentially involves arithmetic reasoning. In Sweeney's most elaborate experiment, large samples of men and women were given a wide variety of problems. Significant sex differences were obtained for groups matched in general intelligence, spatial ability, mechanical comprehension, mathematics achievement, or the amount of training in mathematics. In general, the results confirmed the hypothesis that sex differences favoring men will occur in problems which involve difficulties in restructuring, but not in similar problems which involve no such difficulties.

Milton's study (1957) with college students is pertinent to hypothesis 3 concerning sex differences in problem-solving skills. In this study the Terman-Miles M-F test (1936) was the primary index of sex-role typing, although other M-F questionnaires were also employed. Two types of problem-solving skill, restructuring and straightforward solution, were employed, half requiring numerical solutions and half non-numerical. In general, the results indicate that there is a positive relation between the degree of masculine sex-role typing and problem-solving skill both across sexes and within a sex. When this relation is accounted for, the difference between men and women in problem-solving performance is diminished.

Thus, these studies are consistent with hypothesis 3 that males generally surpass females in problem-solving skills. Moreover, Milton's study also suggests that these differences are accounted for in the typical sex-role development of each sex.

Hypothesis 4, suggesting that males tend to be more concerned with internalized moral standards than females, is consistent with findings in the previously mentioned study of children aged 13 to 18 by Lansky, Crandall, Kagan, and Baker (1961). The *S*s were given a story completion test which was designed to elicit responses regarding severity of moral standards and defenses against guilt following transgression of such standards. *Severity of moral standards* was rated for the degree to which the hero (heroine) punished himself, consciously or unconsciously, for his actions. The boys scored higher than girls on this variable. This finding is considered to support hypothesis 4 in that the *severity of moral standards* was scored when the hero (heroine) punished himself, thus implying that the standards are internalized.

Data from two national sample interview studies of adolescents, reported by Douvan (1957), have relevance here. In answer to two questions to detect self-awareness, boys showed greater concern with establishing satisfactory internal standards and personal control than girls.

Findings in Douvan's studies are also in agreement with hypothesis 5 that females will be more receptive to the standards of others than males. Douvan found that girls are more likely to show an unquestioned acceptance of parental regulation. Koch (1955a, 1955b), along with Sheehy (1938), found girls to be more obedient and amenable to social controls than boys. Thus, these studies seem consistent with hypothesis 5.

In general, the hypotheses that were generated by the theoretical formulation seem consistent with the data. Thus, by postulating that separate learning methods for the two sexes are derived in the process of acquiring appropriate identification, one can formulate hypotheses which are consistent with very diverse findings ranging from the males' superior problem-solving skill to the females' greater need for affiliation. It is not assumed that this formulation, in and of itself, adequately accounts for these diverse findings even though it is generally consistent with them. It is beyond the scope of this paper to attempt to integrate motivation into the formulation, or the psychological implications of anatomical and physiological differences, or adequately to place the role of the father in the development of identification in each sex. These steps, and others, would be necessary adequately to account for these findings. However, it is felt that a formulation along the lines presented here may prove to have a place in more elaborate theories of identification development and may prove helpful in making more sensitive hypotheses concerning psychological sex differences.

SUMMARY

The purpose of this paper is to present a theoretical formulation which postulates basic sex differences in the *nature* of sex-role and parental identification, as well as basic differences in the *process of achieving* such identification. There was a differentiation made between *sex-role identification* and *parental identification.*

The theoretical formulation in this paper used a hypothesis from a previous one (1959) as a postulate from which to derive a number of new hypotheses. That hypothesis suggested that males tend to identify with a cultural stereotype of the masculine role, whereas females tend to identify with aspects of their own mothers' role specifically. For simplicity this paper refers to *masculine-role identification* in males as distinguished from *mother identification* in females.

This formulation adopted the distinction made by Woodworth and Schlosberg (1954) between two kinds of learning tasks, viz., the problem and the lesson. This distinction was used in describing the separate task assigned each sex in learning the appropriate identification. It was further assumed that, in learning the appropriate identification, each sex acquires separate methods of learning which are subsequently applied to learning tasks generally. In learning the mother identification lesson, the little girl acquires a learning method which primarily involves: (a) a personal relationship and (b) imitation rather than restructuring the field and abstracting principles. In solving the masculine-role identification problem, the boy acquires a learning method which primarily involves: (a) finding the goal; (b) restructuring the field; and (c) abstracting principles.

By assuming that these learning methods are applicable to learning tasks generally, the following hypotheses were derived:

1. Females will demonstrate greater need for affiliation than males.
2. Females are more dependent than males on the external context of a perceptual situation and will hesitate to deviate from the given.
3. Males generally surpass females in problem-solving skills.
4. Males tend to be more concerned with internalized moral standards than females.
5. Females tend to be more receptive to the standards of others than males.

These hypotheses were in general agreement with the research findings which were reviewed.

REFERENCES

BACH, G. R. Father-fantasies and father-typing in father-separated children. *Child Develpm.*, 1946, *17,* 63–80.

BROWN, D. G. Sex-role preference in young children. *Psychol. Monogr.,* 1956, *70,* No. 14 (Whole No. 421).

DOUVAN, E. Independence and identity in adolescence. *Children,* 1957, *4,* 186–190.

EDWARDS, A. L. *Edwards Personal Preference Schedule.* New York: Psychological Corp., 1959.

FRENCH, E. G. Development of a measure of complex motivation. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society: a method of assessment and study.* New York: Van Nostrand, 1958. Pp. 242–248.

GESELL, A., *et al. The first five years of life.* New York: Harper, 1940.

GESELL, A., ILG, F. L., *et al. Infant and child in the culture of today.* New York: Harper, 1943.

GOODENOUGH, E. W. Interest in persons as an aspect of sex differences in the early years. *Genet. Psychol. Monogr.,* 1957, *55,* 287–323.

GRAY, S. W., & KLAUS, R. The assessment of parental identification. *Genet. Psychol. Monogr.,* 1956, *54,* 87–109.

HARRIS, D. B. Sex differences in the life problems and interests of adolescents, 1935 and 1957. *Child Develpm.,* 1959, *30,* 453–459.

HARTLEY, R. E. Sex-role pressures and the socialization of the male child. *Psychol. Rep.,* 1959, *5,* 457–468.

HARTUP, W. W. Some correlates of parental imitation in young children. *Child Develpm.,* 1962; *33,* 85–96.

KOCH, H. L. Some personality correlates of sex, sibling position, and sex of sibling among five- and six-year-old children. *Genet. Psychol. Monogr.* 1955, *52,* 3–51. (a)

KOCH, H. L. The relation of certain family constellation characteristics and attitudes of children toward adults. *Child Develpm.*, 1955, *26*, 13–40. (b)

LANSKY, L. M., CRANDALL, V. J., KAGAN, J., & BAKER, C. T. Sex differences in aggression and its correlates in middle-class adolescents. *Child Develpm.*, 1961, *32*, 45–58.

LAZOWICK, L. M. On the nature of identification. *J. abnorm. soc Psychol.*, 1955, *51*, 175–183.

LYNN, D. B. A note on sex differences in the development of masculine and feminine identification. *Psychol. Rev.*, 1959, *66*, 126–135.

LYNN, D. B., & SAWREY, W. L. The effects of father-absence on Norwegian boys and girls. *J. abnorm. soc. Psychol.*, 1959, *59*, 258–261.

MACCOBY, E. E. Role-taking in childhood and its consequences for social learning. *Child Develpm.*, 1959, *30*, 239–252.

McCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

McNEMAR, O. W. Word association, methods of deduction and induction, and reactions to set in good and poor reasoners. *Stanford University Depart. of Psychol., Tech. Rep.* No. 2, 1954.

MILTON, G. A. The effects of sex-role identification upon problem-solving skill. *J. abnorm. soc. Psychol.*, 1957, *55*, 208–212.

SARASON, S. B., DAVIDSON, K. S., LIGHTHALL, F. F., WAITE, R. R., & RUEBUSH, B. K. *Anxiety in elementary school children.* New York: Wiley, 1960.

SEARS, R. R., MACCOBY, E. E., & LEVIN, H. *Patterns of child rearing.* Evanston, Ill.: Row, Peterson, 1957.

SEARS, R. R., PINTLER, M. H. & SEARS, P. S. Effect of father separation on pre-school children's doll play aggression. *Child Develpm.*, 1946, *17*, 219–243.

SEWARD, G. H. *Sex and the social order.* New York: McGraw-Hill, 1946.

SHEEHY, L. M. *A study of preadolescents by means of a personality inventory.* Washington, D. C.: Catholic University of America, 1938.

SHERRIFFS, A. C., & JARRETT, R. F. Sex differences in attitudes about sex differences. *J. Psychol.*, 1953, *35*, 161–168.

SWEENEY, E. J. Sex differences in problem solving. *Stanford University Depart. of Psychol., Tech. Rep.* No. 1, 1953.

SYMONDS, P. M. Life interests and problems of adolescents. *Sch. Rev.*, 1936, *44*, 506–518. (a)

SYMONDS, P. M. Sex differences in the life problems and interests of adolescents. *Sch. & Soc.*, 1936, *43*, 751–752. (b)

TERMAN, L. M., & MILES, C. C. *Sex and personality.* New York: McGraw-Hill, 1936.

WALLACH, M. A., & CARON, A. J. Attribute criteriality and sex-linked conservatism as determinants of psychological similarity. *J. abnorm. soc. Psychol.*, 1959, *59*, 43–50.

WINKLER, J. B. Age trends and sex differences in the wishes, identifications, activities, and fears of children. *Child Develpm.*, 1949, *20*, 191–200.

WITKIN, H. A., LEWIS, H. B., HERTZMAN, M., MACHOVER, K., MEISSNER, P. B., & WAPNER, S. *Personality through perception.* New York: Harper, 1954.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology.* New York: Holt, 1954.

PART II

INFANCY

THE BEGINNINGS OF BEHAVIOR

MATERNAL BEHAVIOR AND ATTITUDES

THE MOTHER-CHILD RELATIONSHIP

THE BEGINNINGS OF BEHAVIOR

10

Research in the Psychological Development of Infants: An Overview

William Kessen

The way in which psychologists see the infant is an important factor in the research on child development. This article contributes to the changing view by pointing up the competent and active infant. Although his focus is on infancy and his research at Yale is with infants, Kessen's conclusions reinforce the thinking on goal-seeking behavior and the role of the environment presented in the articles by Hunt, White, Piaget, and the other authors in Part I.

The infant has not always been treated kindly by American psychologists. Although almost all theories—whether in the tradition of Watson, Freud, or Koffka—celebrate the importance of infant behavior, and claim that the baby is striking proof of the validity of their views, systematic empirical study of the child in his crucial first year has been an on-again, off-again affair. And for a number of reasons, not the least of which is the difficulty of seeing young children in the large numbers that we have at our command in studying the pre-school child or the adolescent. Once a child leaves the hospital after the lying-in period, he is not again easily available for research until he appears in nursery school. It may also be that the infant is so clearly one of us—in that he is human; and so clearly and incomprehensibly different, in that he is a baby—that we have, on occasion, escaped our frustration by constructing theoretical babies instead of observing real ones.

Happily, these disabilities no longer block research. There is evidence, and not only in the United States, that psychologists are studying the infant more closely than ever before. Merely to call the names of investigators and refer to some of their findings would consume many pages. This is not to say, by the way, that the theoretical or constructed child has disappeared. Far from it! Behind each empirical investigation, there is a model, and this model colors and sometimes dominates the interpretation that is given the empirical protocols.

I would like to be able to present a neat, clear (even if artificial) dichotomy or trichotomy of theoretical positions concerning infancy—I recall with some nostalgia our antique friend "maturation *versus* learning"—but the current situation in the psychological study of infancy does not accept such simple classification. It is only a modest exaggeration to say that a recitation of theoretical subtleties would approach the complexity of a recitation of research findings. In the face of this kind of variety, I cannot hope to lay out a complete or even a fair summary of current research and thought about the behavior of the infant. Rather, I will present for your

---

From *Merrill-Palmer Quarterly of Behavior and Development*, 1963, *9*: 83–94.

comment, review, and evaluation, a short set of propositions about babies and studies of babies; under each of these loose-jointed statements, we can examine a part of the research and speculation that has appeared over the last several years.

The first proposition or summary statement that I will propose is that *a comparative psychology of infancy can be anticipated.* Harlow's well-known work (1958)* on affectional systems in the monkey, though incomplete, is as stimulating a body of research as has been done on animal development over the last decade. Less widely known, but of at least equal theoretical impact, are T. C. Schneirla's (1959) speculations about approach and avoidance and their relation to stimulus intensity. Hess (1959), among others, has presented data and commentary on the phenomena of imprinting. Seymour Levine (1957) has contributed a number of papers on the effect of infantile stress on later behavior. These names only begin a list of the researchers who are working on developmental problems with infrahuman animals.

Two general comments are warranted here. The animal work which is now going on in developmental psychology is not "dry-as-dust" laboratory demonstration. Moreover, little of this work leads to procedures routinely applicable to children, in the way that some current studies of reinforcement are; nor is the current animal work aimed at elaboration of the obvious. The psychologists studying animal development are in advance of their colleagues in human developmental studies, not only in regard to novel empirical techniques, but more important, in their willingness to take an intellectual chance or risk a speculation. The second note to be appended to the work with animals is the classical one, namely, the possibility of experimental manipulation of more than a trivial sort. We have only seen the beginning of work with animals, and particularly with primates, that will permit us to examine experimentally propositions that would otherwise remain available only to limited observational examination. Studies of the relation of infant to parent, for example, can be investigated along all relevant dimensions only by the use of animals. This is not to say that once we have found the rhesus we can abandon the human being, but the thoughtfulness and energy of investigators currently working in animal research will have no small impact on current research in the psychology of development.

But let me put aside the allure of precision and control possible with animal work and confine the rest of my general propositions to those about human behavior.

The first proposition about children to be considered, and perhaps the most obvious, is that *infants are various*—young children are different from one another. I may see a straw man when I speak against the notion that human infants at birth, like well-made cigarettes, cannot be distinguished from one another; but there is still abroad in psychology—at least in the academical variety—the feeling that children at birth are, by and large, pretty much undifferentiated protoplasm or no more than merely randomly varying beings. Whatever the present state is of the pure, undifferentiated position in the sociology of knowledge, evidence is accumulating that parents and nurses were right all along—stable differences in behavior can be detected in the first days of life. Hammond (1957) has shown the stability of physical growth patterns. Richmond and his colleagues at Syracuse (1955) have reported psychophysiological stabilities in the newborn. Thomas, Chess, Birch, and Hertzig (1960), although they have published only preliminary reports, have stated that on nine variables—among them, reactivity and irritability—they have found stability in children followed longitudinally over a period of two years. There are some suggestions in Bell's work (1960), and there are some findings in our work on newborns at Yale (1961),* which tend to support a strong generalization that stable individual differences in a large number of behaviors—sucking, general movement, reactivity—exist very early in life. Yet, impressive as it is, the work on the assessment of individual differences among human infants has not, like some of the animal work, been "built out" from novel observations and speculations. Rather, it has come largely from the essential and tedious work of constructing adequate response measures. These advances in technology or method are clear and welcome, but they leave open two larger questions about individual differences.

First, what is the long-range stability or relevance of these differences? It is good to know that the newborn shows stable differences in activity level from his colleague in the next crib, but the importance of this observation is markedly reduced if the difference does not show up in some form later. Among the investigators, other than the Birch group, who have done some interesting speculative work on this score, is the

---

* This article appears in this volume—*Editor's note.*

French psychologist Stambak (1956). She has segregated two groups of infants—hypertonic and hypotonic—and has discussed the relation of this tendency to be active or quiet to such important developmental changes as onset of walking. In addition, the Czech group (Papousek, 1961) is investigating the stability of the infant's behavior during conditioning over the first six months of life. Such studies are provocative curtain-raisers on the intricate question of behavioral stability in infancy.

There is a second question about which we have very little evidence. How are these early behavioral variations related to variations in the environment? How do different combinations of infant and caretaker mesh together? We can tag babies as active or quiet; we can make this discrimination in the first five or ten days of life. We can suspect, too, that some mothers like active babies and some mothers like quiet babies. What do you get when you combine an active baby with a mother who wanted a quiet one or a quiet baby with a mother who wanted an active one? We have very little to go on here, not only because of the obvious technical difficulties of longitudinal studies of this kind, and not only because of the fluidity of our ideas about what is important in the home, but also because, until recently, we have not had reliable ways of describing the young child's environment. The technical advance in the methods of describing newborn behavior have not been matched by methods for describing the home. But here, again, there is promise. Schaefer, Bell, and Bayley (1959) have proposed a parent attitude scale. The important interview work of Sears, *et al.* (1957) provides a framework for the description of parents' behavior. Rheingold (1960)* has recently specified some of the dimensions of variation between home and institution. These papers point the way toward the time when a genuine analysis can be made of the interaction between mother and child. The word "genuine" reflects the hope that this analysis will not be a contaminated one; that we can make assessments of the status of the newborn, independent of observing the mother, and make assessment of the mother, independent of observing the child.

The next summary proposition that I want to suggest warrants detailed examination. I submit that *the young infant is not incompetent* or, by André-Thomas' (1954) catching phrase, "the neonate is not a neophyte."

We have passed the time, not so very long ago, when the newborn was considered to be sensorily bereft (*e.g.*, Preyer's contention that children are born deaf), but the notion of newborn incompetence persists. It has perhaps its strongest statement in the work of the psychoanalysts, especially Spitz (1959), who maintains the existence of a non-differentiated phase in early life, where the newborn does not code inputs at all. In this view of the infant, by no means limited to psychoanalysts, both the baby's sensory capacities and his response capacities are held to be severely limited. The trend of recent research is clearly against this conception of the child. Research on newborn behavior over the last five years has invariably added to the newborn's list of abilities. Peiper (1956) in his encyclopedic treatment, André-Thomas (1954) and his colleagues in Paris, Madame Ste. Anne-Desgassies, and Prechtl (1958) are among the workers who have discussed the extended sensory and response range of the newborn in some detail. Gorman and his associates (1959) have recently found in a study of acuity that the newborn has visual resolving powers which are not markedly inferior to those of the older child. From the research available on the competence of the newborn, let me present three studies in some detail as illustrative and somewhat representative of this newer view of the newborn.

The first study, by Blauvelt (1960), deals with the precision of at least one response the newborn makes. Following up earlier work of Prechtl on head-turning, Blauvelt has studied the baby's response to a very simple stimulation, in which the experimenter moves her finger from the tragus of the baby's ear—the baby lying on its back in the crib—toward the baby's mouth and then away again in a flat elliptical course. It turns out that the baby tracks this movement by turning his head at a speed and to a position that will reduce the distance between his mouth and the stimulating finger. He tracks this movement without special tuition; it is, if you like, built-in. The infant can pick up approaching stimulation and reduce the distance to it very quickly; he can "find" the approaching breast or bottle. What is impressive about this response is the precision of it. This is not the response of a wild newborn, flailing around uselessly and without direction; this is an organism making a precise and exact tracking response. It is a limited skill, to be sure, and certainly not widely generalizable to other activities, but it illustrates the responding precision of some newborn.

---

* This article appears in this volume—*Editor's note.*

The second study illustrative of newborn competence may be one of the most important empirical research products of the last decade in infancy work. Bronshtein, Antonova, Kamenetskaya, Luppova, and Sytova (1958) have described a technique for assessing the limits of sensory differentiation in the infant that promises a precision in psychophysical description that has heretofore been possible only for the much older child. Briefly, the procedure is this. You permit or induce the child to suck, and record his rhythmic response. If, during sucking, you sound a brief tone, say of 512 cycles/sec., the baby stops sucking. When the tone stops, the baby begins to suck again. To a second stimulation of the same tone, he will stop sucking. This sequence can be repeated four or five times for sounds and then when you sound your 512-cycle tone he goes on sucking without interruption. He has adapted to that sound. If, however, you now present a different tone, say one of 1,024 cycles, he will stop sucking. If he continues to suck on the application of the second stimulus, this is presumptive evidence that he cannot discriminate the two stimuli. If he does stop sucking on the second stimulus, if it "undoes the adaptation," then there is evidence that he can discriminate these two stimuli. If this technique is as sensitive as the Russians suggest, we will be able to find out more about the sensory capacities of the young infant than we can find out about the sensory capacities of young five- or six-year-olds. Bronshtein presents data to indicate that the infant makes clearly differential responses to variations in pitch, light intensity, and other stimulus changes. Lipsitt, at Brown, has adapted this technique to a study of olfactory stimulation and has found that not only is sucking inhibited and adapted in this fashion but so also is movement. Just as the Blauvelt study illustrates the possible response flexibility of the newborn, so the Bronshtein and Lipsitt studies indicate the remarkable amount of stimulus coding the newborn is capable of. The world of the infant is not a vast confusing "blob."

Consider yet a third study. In our work at Yale (1963), we have found that if you put a nipple in a baby's mouth, he will stop general movement at once, and when you take it out he will start moving again. This effect appears in the absence of nutrient; the nipple does not supply food—it only provides an opportunity to suck. And, this inhibition of movement takes place in the fourth or third or second, or even first day of life. The child is able to deal with a complex and vitally important input—namely, nipple or sucking—by a very regular response. Nor, apparently, does he have to learn either how to suck or how to quiet. There is of course the argument that he learned the responses *in utero*, but we have hardly advanced beyond Hippocrates' statement of that argument 2,500 years ago.

These studies suggest that the newborn has far greater capacities for sensory discrimination than could have been guessed a decade ago, and though less impressive, the evidence is beginning to indicate that he has surprising response competencies as well. But the evidence for newborn resourcefulness poses a peculiar paradox. To put the question very bluntly, if the human newborn is so capable, why does he not learn more? If he is so capable, why is he so stupid? These questions form the bridge to my next general proposition, one which seems so insecure that I have phrased it in the form of yet another question.

There is early adaptation, but is there early learning? The conflict represented in this question can be expressed simply enough. On the one hand the behavior of baby seems to change over the first few days of life. There are many examples; let me cite just one.

Peiper maintains that there are three techniques of infantile sucking. One of them is the response that most mammals use to get milk out of a breast; it is a lapping response that involves pressing the nipple against the roof of the mouth with the tongue and squeezing milk out of it. Another one is to reduce pressure inside the mouth so as to pull the milk in by a discrepancy in pressure. This is the way most babies suck from bottles. And the third, fairly infrequent technique—confined to bottle-fed babies for obvious reasons—is to bite hard at the back of the nipple and squirt milk into the mouth. This variation is interesting because babies apparently come to use one of these different patterns very quickly. They learn, if "learn" is appropriate, the kind of sucking to use.

The difficulty with calling this kind of change "learning" arises from our failure to demonstrate early learning in a controlled setting. If the newborn is capable of this natural learning it should be possible for a psychologist to teach him something in a systematic learning study. And yet the evidence, controlled evidence for newborn learning, hardly exists. There is research by Marquis (1931), recently replicated in the USSR, showing that the baby adapts to a feeding rhythm, but the evidence does not support the conclusion that learning according to the usual theoretical models takes place in the period of early infancy. The Russians, with their strong demand for environmental control of behavior, have tried a large number of times to condition young infants. Sometimes they are successful; oftentimes they are not. Russian studies do not report conditioning in children under eight or nine days of age, and most conditioning studies indicate that it may take weeks or even months to condition an infant child in the Pavlovian mode (Dashkovskaya, 1953). How do we interpret this curious discrepency between the fact that the human baby seems to adapt his sucking style to his feeding routines on the one hand, and the difficulty that all investigators have had in demonstrating newborn learning on the other?[1]

The following three options seem available to us: First, in spite of my statements about newborn competence, there may be genuine neurological incapacity in the newborn. There is no such thing as early learning, in the usual sense, because the child is not complete. A case for this position can be made. There are data on myelinization, on changes in pattern of EEG, on developments of vision and prehension, on the appearance of smiling—to take the most obvious case—all of which can be used to bolster the view that the young infant is a neurologically deficient organism. Under this reading, how do we account for the changes in behavior that do take place? Perhaps by maintaining that the caretaker becomes more competent. This would be a case of training the parent to adapt more effectively to the child rather than teaching the child to adapt to his environment. And to the data from Bronshtein and Lipsett on the ability of the young infant to make sensory discriminations, we would have to say, "True, infants can make sensory discriminations, but there is no associative coding; there is a deficiency in the hooking of links together."

The second answer, and the one I think that would be given by the learning analysts (Gewirtz, 1961), is that nobody has tackled the problem of early learning. In particular, holders of this position would maintain that the procedures of classical conditioning as used by the Russians are the wrong tactics. What we should do if we want to demonstrate early learning is to use instrumental techniques; that is, to make some effective reinforcement contingent on the occurrence of some response of the infant. For example, let the baby turn his head and then give him something to suck on. This is a testable proposition and it is being tested.

I would like to suggest a third possibility—an unpopular one. In brief, there may be experiential effects that are not learning. To put it another way, not all adaptation of the infant represents either classical conditioning or instrumental learning. I think it is inappropriate to maintain that all changes in behavior that can be related to the child's contact with the environment are the result of reinforcement contingencies. Of course, the instrumental learning position can be made to fit them, but it seems to me that such a forced fit results in theoretical vagueness and a weakening of the instrumental position.

Perhaps in pulling apart the problem of behavioral change in early infancy to exaggerate the variation among options, I have only shown that the resolution of the problem will require revisions in method, new knowledge of infantile neurophysiology, and a reworking of contemporary learning theory.

But consider now another interesting problem which illuminates some theoretical disagreements among students of infancy. Two theoretical positions have occupied this field—the psychoanalytic and the learning-theoretical. Justice can be done to neither in a summary presentation; Rapaport (1959) and Wolff (1960) present the psychoanalytic presuppositions in detail and with force; Gewirtz (1961) has prepared a closely reasoned argument for a learning analysis.[2] Now, there is a new entrant into the field of theories of mother-attachment. John Bowlby in a series of recent papers (1958) has borrowed from the investigators of instinct in animals, a notion that sounds very much like imprinting and has suggested that the child's responses of sucking, clinging, and following lead to mother-attachment. Just sucking, and just clinging, and just following on the part of the child, without obvious reinforcement or redistribution of cathexis, will result in a union between child and mother; much as the chick will imprint on a blinking light. Not only does Bowlby discuss what ties the child to his mother—namely, these three responses—but he also discusses what links the mother to the child. Not only does the child become attached to the mother because of sucking, clinging, and following, but the mother is drawn or attached—Bowlby does not use the word "imprinting"—to the child by the child's smiling and crying. Smiling and crying are held to be congenital or innate releases of maternal behavior.

---

[1] The argument for early adaptation by the infant can probably be made much more forcefully, but a natural history of the first months of life remains to be written.

[2] It should be noted in passing that a learning analysis is both stronger and in a better position for compromise with other views by virtue of a retreat from the drive reduction interpretation of reinforcement and by recent animal studies which show stable secondary reinforcement effects.

It is difficult to evaluate this position and I am hard pressed to invent a satisfactory test for it. Perhaps we must call on animal research to work out the implications of Bowlby's assertions. But the main value of this new view will probably be the value of all theories of development—that they jog thinking, they make people run a study just to see what happens. Certainly Bowlby's ideas have had that effect. His own research with Robertson on separation (1952), the work done by Schaffer (1959) on hospitalization of young children, and an unpublished study by Ainsworth (1961) have demonstrated the provocative effect of these speculations. One of the achievements of the work done by Ainsworth, in Uganda, is that, instead of discussing mother attachment as a unitary notion, she has, in these longitudinal field observations of the child between four weeks of age and fourteen months, described some ten or twelve indexes of the child's attachment to the mother, and in this way has made possible a more subtle analysis of the relation than we have had heretofore.

It is interesting to note, as an adjunct to the problem of mother-attachment, that something very curious indeed seems to happen to children near the middle of the first year. Ambrose's (1961) results on smiling indicate that at 17–25 weeks, general social smiling begins to decay and the child begins to smile only at its caretakers. Schaffer's work indicates that children who are hospitalized before they are 28 weeks old, accept hospital routines and separation easily; children hospitalized after 28 weeks-of-age show striking symptom patterns of distress and refusal to accept normal hospital care. Ainsworth finds that almost all of her criteria of mother-attachment begin to show transition in the period from 17 to 30 weeks, with much of the change occurring in the narrow band between 25 and 28 weeks. Somewhere in the middle of the first year, the child appears to shift from being attached to human beings at large to being attached to one, or two, or three human beings.

The Ainsworth study is comparable in its impact to Rheingold's (1956) study of caretaking in institutionalized infants—both of these studies represent the payoff for the theoretical positions underlying them. The psychology of infancy undoubtedly profits from being in a state of theoretical dis-equilibrium, and the diversity of ideas about the nature of the child's attachment to his mother will almost certainly be productive of important empirical advances.

Consider one last generalization about infancy. It is one where contention, compromise, and reciprocation among theoretical positions has already resulted in general agreement. *The infant is active, and the relation of infant and caretaker is reciprocal.*

It is on this issue that the psychologist's view of the child has changed most dramatically in recent years. The model of the child which was drawn from Pavlov through Watson, and supported by the development of learning psychology in the United States, was of a recipient organism—a reactive one. Behavior at any particular time is the function of the current stimulating environment. This remains technically a sound view, but the effect of it on the psychology of the infant was to diminish our appreciation of how complicated and subtle is the child.

Not only can the child be usefully seen as active, rather than merely as reactive, but it may also be useful to think of even the infant as a problem-solver. Certainly the child, like the adult, can be seen as encountering problems in his environment. At least from the age of six months, the child's behavior can be discussed in terms of discrepancy, goal-seeking, means to an end, and so on. One student of children has not deviated from this view of the active searching child. Piaget and his students have seen the child, especially the infant, as being in a constant exchange with the environment, meeting its demands, and what American investigators somehow forgot, making its own demands on that environment.

The shift in point-of-view—to set the antithesis sharply—has been from the child who is a passive receptacle, into which learning and maturation pour knowledge and skills and affects until he is full, to the child as a complex, competent organism who, by acting on the environment and being acted on in turn, develops more elaborated and balanced ways of dealing with discrepancy, conflict, and dis-equilibrium. This shift, I believe, is of incalculable implication and seems to have been accepted to some degree by almost all students of children. Bowlby emphasizes the control by the child in crying and smiling; psychoanalytic theory makes more space for autonomous ego functions; child psychologists dedicated to a learning analysis speak of the child as active; and I suspect Piaget thinks of how he knew it all the time. But this shift only sets the problem for the psychology of the infant; questions abound. What is a "problem" for the infant? What is an environmental discrepancy for the newborn, for the six-month old, for a walker? Do Piaget's speculations about assimilation, accommodation, and equilibration have more than a metaphorical value? Can child psychologists follow the lead of psychologists of cognition in adults, who use computer analogies? Can we build a theory of cognitive development without the use of terms like reinforcement,drive, or dissonance resolution?

Only one thing seems certain. We are better equipped, with attitude and technique, to make a systematic and meaningful analysis of infant behavior than ever before. The current psychology of infant behavior, by and large, is managing to steer skillfully between the Scylla of "Oh, Oh, look what the baby did!" and the Charybdis of "But the theory says thus and so." We are engaging in hot, theoretical debate, but more and more the debate refers back to the child—back to the theory-illuminated facts.

REFERENCES

AINSWORTH, M. D. The development of infant-child interaction among the Ganda. Paper read at Tavistock Study Group on Mother-Infant Interaction, London, 1961.

AMBROSE, J. A. The development of smiling response in early infancy. In B. M. Foss (Ed.), *Determinants of infant behavior.* New York: Wiley, 1961.

ANDRÉ-THOMAS Ontogénèse de la vie psycho-affective et de la douleur. *Encéphale,* 1954, *43,* 289–311.

BELL, R. Q. Relations between behavior manifestations in the human neonate. *Child Develpm.,* 1960, *31,* 463–477.

BLAUVELT, H., & McKENNA, J. Capacity of the human newborn for mother-infant interaction: II. The temporal dimensions of a neonate response. *Psychiat. Res. Rep.,* 1960, *13,* 128–147.

BOWLBY, J. The nature of the child's tie to his mother. *Int. J. Psychoanal.,* 1958, *39,* 1–24.

BRONSHTEIN, A. I., ANTONOVA, T. G., KAMENETSKAYA, A. G., LUPPOVA, N. N., & SYTOVA, V. A. On the development of the functions of analyzers in infants and some animals at the early stage of ontogenesis. In *Problems of evolution of physiological functions.* OTS Report No. 50-61066. Translation obtainable from U.S. Dept. of Commerce. Moscow: Acad. Sci., 1958.

DASHKOVSKAYA, V. S. First conditioned reactions in newly born children in normal state and in certain pathological states. *Zh. vyssh. nervn. Deiatel.,* 1953, *3* (2), 247–259.

GEWIRTZ, J. L. A learning analysis of the effects of normal stimulation, privation, and deprivation on the acquisition of social motivation and attachment. In B. M. Foss (Ed.) *Determinants of infant behavior.* New York: Wiley, 1961.

GORMAN, J. J., COGAN, D. G., & GELLIS, S. S. A device for testing visual acuity in infants. *Sight-Saving Rev.,* 1959, *29,* 80–84.

HAMMOND, W. H. The constancy of physical types as determined by factorial analysis. *Hum. Biol.,* 1957, *29,* 40–61.

HARLOW, H. F. The nature of love. *Amer. Psychologist,* 1958, *13,* 673–685.

HESS, E. H. Imprinting. *Science,* 1959, *130,* 133–141.

KESSEN, W., & LEUTZENDORFF, A. M. The effect of non-nutritive sucking on movement in the human newborn. *J. comp. physiol. Psychol.,* 1963, *56* (1), 69–72.

KESSEN, W., WILLIAMS, E. J., & WILLIAMS, J. P. Selection and test of response measures in the study of the human newborn. *Child Develpm.,* 1961, *32,* 7–24.

LEVINE, S. Infantile experience and resistance to psychological stress. *Science,* 1957, *126,* 405.

MARQUIS, D. P. Can conditioned responses be established in the newborn infant? *J. genet. Psychol.,* 1931, *39,* 479–492.

PAPOUŠEK, H. A physiological view of early ontogenesis of so-called voluntary movements. In P. Sobotka (Ed.), *Functional and metabolic development of the central nervous system.* Prague: State Pedagogic Publ., 1961.

PEIPER, A. *Die Eigenart der Kindlichen Hirntätigkeit,* 2nd ed. Leipzig: Thieme, 1956.

PRECHTL, H. F. R. The directed head turning response and allied movements of the human baby, *Behaviour,* 1958, *13,* 212–242.

RAPAPORT, D. The structure of psychoanalytic theory: A systematizing attempt. In S. Koch (Ed.), *Psychology: A study of a science,* Vol. 3. New York: McGraw-Hill, 1959.

RHEINGOLD, H. L. The modification of social responsiveness in institutional babies. *Monogr. Soc. Res. Child Develpm.,* 1956, *21* (2).

RHEINGOLD, H. L. The measurement of maternal care. *Child Develpm.,* 1960, *31,* 565–575.

RICHMOND, J. B., & LUSTMAN, S. L. Autonomic function in the neonate: I. Implications for psychosomatic theory. *Psychosom. Med.,* 1955, *17,* 269–275.

ROBERTSON, J., & BOWLBY, J. Responses of young children to separation from their mothers. *Courrier de la Centre Internationale de l'Enfance,* 1952, *2,* 131–142.

SCHAEFER, E. S., BELL, R. Q., & BAYLEY, N. Development of a maternal behavior research instrument. *J. genet. Psychol.*, 1959, *95*, 83–104.

SCHAFFER, H. R., & CALLENDER, W. M. Psychologic efforts of hospitalization in infancy. *Pediatrics*, 1959, *24*, 528–539.

SCHNEIRLA, T. R. An evolutionary and developmental theory of biphasic processes underlying approach and withdrawal. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1959.* Lincoln: University of Nebraska Press, 1959.

SEARS, R. R., MACCOBY, E. E., & LEVIN, H. *Patterns of child rearing.* Evanston, Ill.: Row Peterson, 1957.

SPITZ, R. A. *A genetic field theory of ego formation; its implications for pathology.* New York: International Universities Press, 1959.

STAMBAK, M. Contribution à l'étude du développement moteur chez le nourrisson. *Enfance*, 1956, *9*(4), 49–59.

THOMAS, A., CHESS, S., BIRCH, H., & HERTZIG, M. E. A longitudinal study of primary reaction patterns in children. *Comprehensive Psychiat.*, 1960, *1*, 103–112.

WOLFF, P. H. *The developmental psychologies of Jean Piaget and psychoanalysis.* New York: International Universities Press, 1960.

11

Selection and Test of Response Measures in the Study of the Human Newborn

William Kessen, E. Jane Williams, and Joanna P. Williams

One of the problems in the careful study of infants has been the refinement of research procedures. In this paper Kessen and his colleagues describe the use of photography along with trained observers for the recording of data on newborns. In addition to this research technique, they present both conclusions about the individuality of these babies and questions about the psychological meaning of such individuality.

Thirty years ago, Weiss stated the problems of response selection in studies of the human newborn:

> . . . There is very little coordination of the different movements. . . . The assumption that no movement is ever *repeated exactly* becomes a very obvious fact. It seems at first futile to speak of classifying the behavior. While this is true theoretically, as a matter of fact we do group together certain types of movements according to their resemblances (1929, p. 454).

Unfortunately, Weiss's criterion of classification —the occurrence of "resemblances"—does not provide an unambiguous rule for ordering the behavior of newborns. Should we attend to resemblances in anatomical topography, or speed, or temporal overlap, or effect on the environment, or relation to antecedent stimulation? In fact, all of these criteria for classification have been used by students of neonatal behavior, with little evidence that a simple general rule for response definition is emerging. The problem of response definition is by no means confined to research with infants—it has long been an issue in general psychology—but the newborn presents special difficulties in answering this most basic of research questions: What are the effective response-classes?

In the observation of animals and older human beings, psychologists of all theoretical commitments depend on *movement in relation to the environment* and on *speech* to provide the most elemental definition of response-categories. Whether in the study of maze-learning or of psychotherapy, the psychologist of mature behavior can make use of the fact that his subjects move through space or speak; the student of newborns can claim no such advantage. This difference turns out to be quite important in the first stages of the search for meaningful response-categories. Although general definitions of response-classes have been developed (e.g., those responses which increase in probability of occurrence with the occurrence of reinforcement), a great deal of psychological observation still depends on the use of categories common to "natural" everyday usage. Thus, the behavior of picking up a cigaret and the behavior of picking up a rattle can "naturally" be put in the response-class Picking Up, even though the responses may differ along many dimensions. Similarly, the expression "I hate dishwashing" and the expression "Golf is repulsive to me" will naturally indicate the same response-class of Disliked Activities even though they are phoneti-

---

From *Child Development*, 1961, *32*, 7–24. Reprinted with permission of the Society for Research in Child Development.

cally quite different and may differ on other grounds as well. People organized the activities of other people into response-clusters long before there were professional psychologists, and these vernacular rules for the classification of behavior seem to depend in large measure on characteristics of language and on regularities in response to objects (including other organisms) in the environment. When the psychologist sets out on his systematic study, he almost invariably adapts his observation to these everyday classificatory rules. Such a strategy is reasonably successful when used with mature human beings and even with animals, but in most cases of newborn behavior the "natural" rules are inapplicable. It is largely for this reason that the behavior of newborns is seen as meaningless; we are unable to organize the neonate's behavior with the rules of classification natural for older organisms.[1]

Curiously enough, the lack of obvious criteria of meaningfulness in newborn behavior also precludes exclusion of response-classes; that is, any behavior might be productively studied. This ambiguity, which is a problem for all psychologists but of special importance for those of us interested in early behavior, forces attention to the more general question: What do you want the response-categories for? There are several ways in which this question has been answered in the past, and a brief review of some of the positions taken will put the issue of response selection in a sharper light.

In the tradition of comparative psychology, the human newborn can be seen as a separate species, or at least a separate breed, of interest to the psychologist because of his variation from the behavior of other animals. Much of the classical research on neonatal behavior fits with this view of the infant. Can he discriminate color? What are his taste sensitivities? Can he learn? Is he responsive to privation? What reflexive behavior does he show? Questions of this order, which led to the accumulation of important observations of infant behavior, were almost invariably derived from general psychology; those independent variables were manipulated which had been important in the study of other animals, and response definitions were developed which showed more-or-less reliable relationships with external stimulation. Dennis's (1934) survey of the response repertory of the human infant lists the response-classes which resulted from this pioneer work on neonatal reflexology, and his summary is convincing demonstration that a relatively large number of response categories for neonatal behavior can be described under the criterion of close temporal relation to external stimulation.

In the course of early research on newborn behavior, there was developed a response-class very different from the simple "reflexive" activities which dominate Dennis's list. Called variously "mass activity" or "general activity," this response category was the antithesis of the stimulus-response descriptions and had as its criterion of inclusion *any* detectable movement of the newborn. Several techniques were used to determine measures for this response-class, but in general they aimed at representing all movements of the infant, whether these movements, collected in smaller subclasses, could be reliably related to external stimulation or not. For example, a response class defined in this way was the chief tool in Irwin's fundamental work describing the effect of several antecedent manipulations on neonatal activity (e.g., 1930). There was promise, during the thirties, that "general activity" would serve for the study of newborn behavior the same broad function that "pressing a bar" or "running a maze" had served for other psychological problems. Perhaps because the specific measures devised for neonatal activity were not sensitive enough to detect response changes of brief duration, this promise was not fulfilled.

The study of the newborn as species, directed more by interest in reliable description than by commitment to elaborated theoretical issues, produced a natural history of neonatal behavior which was fundamental to later study and developed a response class—general activity—which was shown to be responsive to manipulation.

Running alongside research interest in newborns as behavioral variants, and never independent of it, has been the study of neonatal behavior as the precursor of later behavior. In this view, the newborn is seen as part of a long development toward adult human functioning, and response-classes have been sought which will show continuity of this development over time. However, as one might expect from our earlier discussion, there are no obvious and natural continuities between the behavior of newborns and the behavior of older human beings. As a result, the search for genotypic continuity in the face of phenotypic diversity has led to a stronger emphasis on theory—that is, on postulated continuity—than was the case for the comparative psychologists of infancy. From this search for developmental laws have come the three most important contributions to a general theory of newborn behavior—Freud's theory of personality development (e.g., 1905), Gesell's theory of maturation (e.g., 1945), and Piaget's theory of the growth of intelligence (e.g., 1952). Each of these systems, devised to deal with the entire range of human development, makes statements about the important response-classes to be observed in the study of newborns.

---

[1] Wolf (1957) has suggested that the basic natural rule for organizing the behavior of human beings depends on the observation of "goal" or "endpoint" behavior. If this is the case, then the apparent meaninglessness of the newborn's activities may represent the inability of adult observers to recognize the newborn's "goals."

Gesell's work has had by far the widest research impact, leading to a line of studies on the normative development of the human being, particularly with regard to relatively simple motor behavior. Piaget's interest has been in aspects of neonatal behavior which show the operation of what he considers general intellective functions, and his observations (1952) have as a theme the infant's development of increasingly effective ways of dealing with the environment. To date, no systematic study has been made of newborn behavior with Piaget's postulations as guide. The developmental theory outlined by Freud, which emphasized the early expression and control of impulse, has produced a great deal of commentary and some provocative research (e.g., Spitz & Wolf, 1949). Its implications for the study of neonatal behavior have been treated in detail by Lustman (1956), who also discusses the data on newborn behavior relevant to psychoanalytic theory.

Whatever theoretical position is taken, the study of the immature organism as precursor of the mature organism involves two interrelated problems: the establishment of developmental sequences common to all human beings (e.g., changes in prone progression, psychosexual development) and the determination of parameters which permit the description of individual differences (e.g., accelerated development, regression rate). In the specific instance of the newborn, one may ask: What behavior or change in behavior is common to all infants (e.g., occurrence of sucking movements when orally stimulated). Or one may ask: What behavior or change in behavior is characteristic of one child or of a limited group of children (e.g., variations in intensity of sucking movements). . . .

In short, the consideration of the newborn as a precursor brings to the problem of response selection two points of emphasis: the relevance of neonatal behavior to later behavior and the detection of individual differences. Attention to these aspects of the problem also makes clear that the lack of obvious and common-sense continuities between neonatal and later behavior requires that the selection of response-classes be directed in the first instance by hunch or preconception or theory.

Once a response-class has been selected for study, however, whether on the basis of a specific developmental theory or on the basis of a less well-organized hunch, criteria of empirical relevance become important, and the usefulness of a response-class in the systematic study of newborn behavior comes to be determined by factual considerations. Can the response be observed with high intersubjective agreement? If it is held to indicate individual differences, are values of the response stable for a neonate from occasion to occasion? Is the response related in a systematic way to age, antecedent conditions, or other responses?

The problem of classification posed by Weiss goes beyond the statement of "resemblances," both because the resemblances are not always obvious and because the resemblance selected may not meet empirical requirements for a meaningful response category. The selection of response-classes in the study of newborn behavior will be at first directed by hunch or theory, and those response-classes which are shown to be reliable, stable, and systematically related to other events will survive as useful parts of a system of newborn behavior.

The research to be reported in part here grew from an interest in the early history of tension and tension-control. As Wolf has pointed out, the history of the child can be seen as the history of his responses to developmental "problems," and "the earliest behavioral manifestations of problem-solving are the baby's modes of tension reduction" (Gardner, 1955, p. 507). A closely related concern with tension or disequilibrium as important in the understanding of behavior is reflected in the notions of "drive" in contemporary learning theory and of "instinct" in psychoanalytic theory. In the apparent chaos of neonatal behavior, two primitive generalizations related to these theoretical notions seem to be warranted. First, there are changes from occasion to occasion in the tension or disturbance or disequilibrium of the human newborn. He may be quietly asleep and the next moment awake and screaming; he may be slowly moving his arms and legs and then suddenly startle; he may produce sighs of contentment in the observer or twitches of distress. The second generalization about which we can be relatively certain is that these early variations in tension come to be more and more closely related to environmental events. In fact, to the degree that one is able to set up a stable relation between some aspect of the environment and changes in the infant's level of tension, to that degree will his behavior become meaningful. The infant may become quiet when lifted to the feeding position, for example, or begin to cry "for food" at regular temporal intervals. These basic observations, coupled with the theoretical importance of tracing the early history of tension, led to the first guiding question of the present research: What variables or events influence variation in neonatal tension and tension-control, and what is the early developmental course of this influence?[2]

Beyond variations in tension and tension-control in the same children from occasion to occasion, there are theoretical and suggestive empirical grounds for believing that reliable variation in tension occurs between subjects, that is, that tension is a source or index of individual difference. Moreover, in the case of developing tension-control, there is reason to believe that different infants may show different tension-control patterns, some children being more responsive to tactile stimulation, others to gastrointestinal changes, and so on. From these considerations emerged the second general question of the research: Can stable individual differences in patterns of tension and tension-control be described for the newborn?

Clearly, these most general statements of the problem do not provide obvious and natural response measures; there remains a wide range of observational protocols which might serve as indicators of tension, and there are perhaps even more variables to be investigated as important in tension-control. For the first empirical analysis of the problem, therefore, the study was narrowed to an investigation of general movement and crying as indicators of tension and of mouthing and hand-mouth contacting as potential tension-control activities. The choice of crying was based on the almost universal popular generalization that it is an expression of distress by the infant; it most assuredly distresses the adult observer. The use of general movement as a partial definition of tension is also justified in common-sense psychology—"a quiet baby is a happy baby"—but it receives support from other sources as well. General movement or activity has been ascribed substantial theoretical importance in an understanding of neonatal functioning; as mentioned earlier, measures of movement have had empirical use of some success; and perhaps of greatest relevance to long-term research goals, we hoped to develop a measure of movement which would be sensitive to changes of short duration in the behavior of the newborn.

The selection of oral behaviors as possibly relevant to tension-control was dictated by both observation and theory. Whatever other disagreements may separate theorists of early development, all agree that mouthing and sucking are fundamentally important responses in the newborn. Moreover, preliminary observation had suggested that two oral responses—hand-mouth contacting and mouthing—could be reliably observed and might reflect systematic individual differences.

Thus, the larger questions raised earlier become reduced to the more readily testable problem: What events or variables control variation in crying, general movement, mouthing, and hand-mouth contacting in the neonate, and how does this control change during the first days of life? There are many routes along which this question can lead, but our initial step was to safeguard more complicated issues by first determining whether the measures chosen meet the empirical criteria stated earlier, that is, by assessing the reliability, stability, and relatedness of our measures. Such is the aim of the present report. Can general movement, crying, and aspects of oral behavior be measured reliably and sensitively? Do these behaviors show consistent individual differences? Do these behaviors show systematic relations with each other and with situational variables?

METHOD

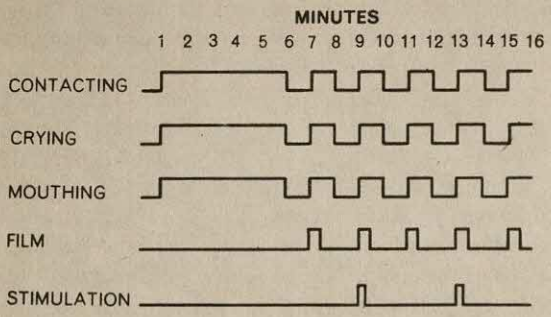
Subjects

The subjects of the research were mature, normal infants in the maternity service of Grace-New Haven Community Hospital. Each subject was seen several times during his hospital stay; most subjects were observed at least once each day (*see* below, *Schedule of Observations*). In order

---

[2] There are two other closely related ways in which early changes in human behavior can be discussed. The child may be seen as increasingly *in control of his environment*, that is, as becoming more competent. Alternatively, the child may be seen as coming to be increasingly *under the control of his environment*, that is, as becoming more predictable in relation to external events.

Fig. 1: *Schedule of observational procedures, group IIa.*



to eliminate systematic bias in the selection of subjects, the following criterion was established for inclusion in the sample: that, whenever we began an observation series, the subject chosen be the youngest newborn in the nursery who had not yet been fed. Thus, the age of subjects at the time of their first observation ranged from 1½ hours to 13½ hours, with the median age at 8½ hours. This criterion of selection had the further advantage of permitting the comparison of neonatal behavior before and after the first feeding.

The present report is based on the repeated observation of 50 newborns. The specific observational procedures used were changed as the research became more focused and detailed, with the subjects falling into two main groups: group I ($N = 34$), where our chief interest centered on the observation of hand-mouth contacting; and group II ($N = 16$), where a detailed examination was made of mouthing, crying, contacting, and movement. Each of these groups was subdivided to provide information on special problems (*see* below, *Schedule of Observations*).[3]

Schedule of Observations

Typically, each infant was observed on each of the five days of his hospital stay. Shortly before noon, the infant was brought from nursery to the observation room, which was an unoccupied nursery on the same floor of the hospital. Blankets and other wrappings were removed, and the newborn (dressed in hospital shirt and diaper) was placed supine on the observation pad. If the child was crying, he was held or rocked in arms until he quieted. Once the observation was begun, however, the child was not handled except in cases of acute distress, when a nurse was called to tend to him. The observation period covered from 10 to 30 minutes, depending on the group to which the child was assigned, after which the child was returned to nursery. The schedule of observations for each group of subjects was as follows:

Group Ia ($N = 25$): General observation, followed by five-minute observation (checklist) by at least two observers of hand-mouth contacting. This procedure was repeated, when possible, on days 1, 2, 3, 4, and 5. For 10 subjects in this group, a pilot run of motion photography (30 seconds in duration) was made on three successive days.

Group Ib ($N = 9$): General observation, followed by five-minute observation (checklist) by at least two observers of hand-mouth contacting. This procedure was repeated, when possible, three times each day (at 8–9 a.m., at 11–12 a.m., and at 5–6 p.m.) on days 1, 2, 3, 4, and 5.

Group IIa ($N = 11$): General observation, followed by five-minute observation (key-depression with record made on an Esterline-Angus operations recorder) of crying, mouthing, and hand-mouth contacting. After this observation, the same measures were taken over five one-minute observation periods, each separated from the next by one minute of no recorded observations. During each of the one-minute observation periods, a 30-second sample of motion photography was obtained. For 10 to 15 seconds of minute-observations 8 and 12, the observer inserted his shielded index finger into the baby's mouth. Figure 1 shows this procedure in diagrammatic form. The procedure was repeated on days 1, 2, 3, 4, and 5.

Group IIb ($N = 5$): Procedure was exactly the same as that for group IIa, except that no oral stimulation was applied and observations were not made during minutes 8, 10, 12.

Response Measures

Hand-mouth contacting. The newborn's response of bringing his fingers or hands to his mouth was recorded in two ways. The first, used with group I, involved placing a checkmark on a mimeographed sheet whenever a contact occurred, recording thereby the finger which made contact and the *frequency* of contacts over a predetermined period of time. For group II, the depression of one of two telegraph keys marked on a line on an Esterline-Angus operations recorder (Recorder), indicating which hand made contact, the *frequency* of contacts over a predetermined period, and the *duration* of hand-mouth contacting.

---

[3] The limitations of the clinical setting in which the research was conducted and occasional apparatus failures led to the discarding of some data on some subjects. As a result, in the analyses presented later, Ns will vary from those given here.

Mouthing. In addition to the movements of the newborn's mouth which can be observed in the feeding situation and which frequently accompany hand-mouth contacting, mouthing movements occur in the newborn in the absence of external stimulation. For group II, a telegraph key was depressed, marking a line on the Recorder whenever the infant's lips moved in a characteristic sucking pattern, either in the presence or absence of external stimulation. In the analysis presented below, the *duration* of mouthing is the measure of this activity.

Crying. For group II, a telegraph key was depressed, marking a line on the Recorder whenever the infant made an audible cry. In order to avoid observer anticipation of crying, the recording key was depressed only on the exhalation part of the crying cycle. As in the case of mouthing, *duration* of crying is the measure used for analysis.

For all duration measures, the Recorder tape was scored by seconds of observation, and a score of 1 was assigned for each second in which hand-mouth contacting, mouthing, or crying was recorded. Thus, for all three duration measures, an infant's score could vary from zero to whatever number of seconds elapsed during the observations. For comparability among observations of different lengths, all scores were reduced to a "per 30-seconds" base. Therefore, for any observation, an infant's crying, mouthing, or hand-mouth contacting duration score varied in a range from zero to 30.

General movement. Each child in group II and 10 children in group I were photographed by a fixed-position overhead camera for one or more runs of 30-seconds duration on several days of the lying-in period. The whir of the camera's spring drive did not, by and large, alter the infant's behavior. No artificial illumination was used other than the overhead lights permanently installed in the nursery. These filmstrips were analyzed to provide a measure of the horizontal displacement of hands and feet over five-second intervals. The details of the technique are described elsewhere (Kessen, Hendry, & Leutzendorff, 1961); although the technique is tedious and time-consuming (requiring the examination and measured comparison of 50 film-frames per child per observation), it provides a reproducible and permanent record of the infant's movements.

RESULTS

Reliability of Measures

Hand-mouth contacting. The reliability of the checklist technique was high; correlations between pairs of observers were invariably above .90, and, for the two observers who did the bulk of observations for group Ia, the interobserver correlation was .96. The Recorder technique used for group II showed similar intersubjective reliability, with the protocols of the two most frequent observers showing a correlation of .96.

Mouthing. The occurrence of mouthing was the event most difficult to discriminate in the study. Nonetheless, test runs for reliability produced an interobserver correlation coefficient of .93.

Crying. Although there is some ambiguity in the observation of crying (e.g., unvoiced cries, "crying face" without sound), observers did not have the reservations typical for the observation of mouthing, and test runs for reliability showed an interobserver correlation of .99.

Movement. The assessment of reliability for the movement measure was made by comparing the movement values assigned to children by two analyzers after they had independently sketched and measured the appropriate frames of film. The discriminations required for this task were matching one picture against the next and measuring linear displacement of limbs with a millimeter rule.[4] The correlations which resulted from this assessment ran above .96.

The measures set up for general movement, crying, mouthing, and hand-mouth contacting were shown to be very highly reliable; moreover, the absolute values recorded by different observers did not show more than negligible variation in means and variances.

Individual Stability of Measures

Being assured that reliable protocols can be established for movement, crying, hand-mouth contacting, and mouthing, we put this next question to the data: Are these measures stable from infant to infant over the lying-in period; that is, is there evidence of consistent individual variation among newborns for these response categories?

---

[4] See Kessen, Hendry, and Leutzendorff (1961) for further details.

Hand-mouth contacting. Hand-mouth contacting is by no means an infrequent response in the human newborn. Without exception, every infant seen in pretest series and in the research proper was observed to bring his hand to his mouth at least once during the five-minute observation periods. Moreover, the variation in frequency of the response is great; among the 32 subjects who were observed on each of the five days of hospital stay, the average number of hand-mouth contacts during five minutes ranged from 3 to above 30, with two subjects making contact 60 times in a single observation period.

In group I (checklist group), 16 newborns were seen on all five days at the midday observation period; this provides an array of 80 contacting scores which can be evaluated in two ways to indicate individual stability. The first of these is a rank order procedure, in which frequency of hand-mouth contacting is ranked for each day, and the coefficient of concordance computed for the resulting array of ranks. This statistic had a value of .51 ($p < .01$); indicating that a significant proportion of variation in ranked frequency of hand-mouth contacting could be assigned to individual differences. The same finding can be shown in an analysis of variance of the day-by-subject array of hand-mouth contacts; such an analysis shows a significant variance ($F = 3.47$, $df = 15/60$, $p < .01$) ascribable to between subject variation.

Group II (Recorder) provided an opportunity to check the replicability of this finding. Sixteen subjects were seen on five days of hospital stay, and the resulting coefficient of concordance was once again significantly different from chance ($W = .34$, $p < .05$). The analysis of variance of raw contacting scores showed a parallel outcome ($F = 2.19$, $df = 15/60$, $p < .05$).

In evaluating these findings, it must be remembered that they are based on a relatively brief sample of the infant's total stay in the hospital (about 0.5 per cent), and under circumstances (e.g., variation in maternal handling) which introduce considerable unsystematic variation. Therefore, the stability demonstrated here is almost certainly an underestimate of the stability to be expected from a longer time sample or in an artificially controlled setting.

It is interesting to note that the data from group II provided no evidence of significant individual stability for *total duration* of hand-mouth contacting ($W = .22$) nor for *mean duration per contact* ($W = .10$). This result suggests that getting the hand to the mouth may be under the control of variables different from those which control the maintenance of a contact once made.

Mouthing. Some activity of the mouth, whether responsive to external stimulation or not, is a strongly prepotent response in the human newborn. Among the 16 infants in group II, the five-day average duration of mouthing during a five-minute observation ranged from 56.4 seconds (18.8 per cent of time observed) to 189.6 seconds (63.2 per cent) with a mean at 112.5 seconds (37.5 per cent). For this measure, a coefficient of concordance computed on an array of ranks does not reach statistical significance ($W = .21$); the corresponding analysis of variance shows a marginally significant variance attributable to between subject variation ($F = 2.09$, $df = 15/60$, $p < .05$). For the conditions of observations obtaining in the present study, mouthing is apparently not as sensitive a measure of stable individual variation as is hand-mouth contacting.

Crying. For the 16 subjects in group II, the crying measure showed wide variation, with mean duration per five-minute observation over the lying-in period ranging from 6.8 seconds (2.3 per cent of time observed) to 117.6 seconds (39.2 per cent) with a mean at 54.5 seconds (18.2 per cent). The distribution of crying durations was irregular in shape, with 27 of 80 observations showing no crying at all. No significant individual differences were found on the crying measure. General observation of the infants suggested that this response is both too infrequent and too much under the control of fluctuating external stimulation (e.g., wet diapers) to be a convincing index of individual variation.

Movement. The 30-second filmstrips run for each subject in group II on each day of his hospital stay provided a measure of neonatal movement. The sum of displacements over five-second intervals of right and left palm and right and left ankle was the primary score submitted to analysis. For 10 subjects in group IIa, this score was computed for five daily observations and these scores combined to provide a single daily value of movement for each infant. The resulting coefficient of concordance was .53 ($p < .01$), confirming the general observation that a significant portion of the variation in newborn movement can be assigned to individual variation. The corresponding analysis of variance shows the same result ($F = 3.86$, $df = 9/36$, $p < .01$). It should be noted, too, that this score combines observations of "spontaneous" movement with observations of movement following external stimulation (i.e., insertion of the observer's shielded finger in the infant's mouth at observation minutes 8 and 12 each day). Individual stabilities were clearly potent enough to override the effect of this intervention, and a high degree of confidence can be placed on the finding of consistent subject differences.

This stability in neonatal movement can be seen even more dramatically by examining the scores of groups IIa and IIb for the *first* 30-second observation each day. The coefficient of concordance for 15 subjects was .49 ($p < .01$); the analysis of variance showed significant variation between subjects ($F = 3.33$, $df = 14/56$, $p < .01$). Thus, a stable measure of individual variation in newborn movement can be computed, *based on a 30-second sample* on each day of hospital stay.

The range of variation among neonates in movement over a five-second measured interval, averaged for the lying-in period (groups IIa and IIb, first filmed observation each day) was from 1.50 cm./limb to 5.70 cm./limb, with a mean at 3.65 cm./limb.

In assessing the four measures under examination in the present study for their sensitivity in showing stable individual variation, we may conclude that hand-mouth contacting and general movement vary in a significant way from child to child, that mouthing is a marginally sensitive measure of individual variation, and that crying as measured in the present setting does not provide a reliable index of individual variation.

Relatedness of Measures

Day-to-day effects. During the first five postnatal days, the newborn does not merely age; he is trained to a new feeding technique, he is extensively handled, if he is a male he undergoes minor surgery, and so on. It is therefore difficult to assign sources of variation in day-to-day effects. However, several meaningful questions may be asked. Does the behavior show regular and systematic change over days? Are there significant differences between day 1 (when, for our sample, no feeding had occurred) and later days? Do relationships among variables change over time? There are 15 subjects in group II on whom complete data are available which can be used to answer these questions.

Fig. 2: *Changes in hand-mouth contacting, mouthing, and crying over the first five days of life.*

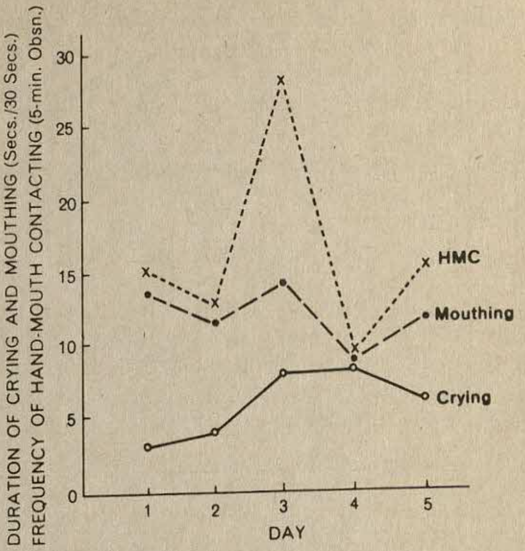
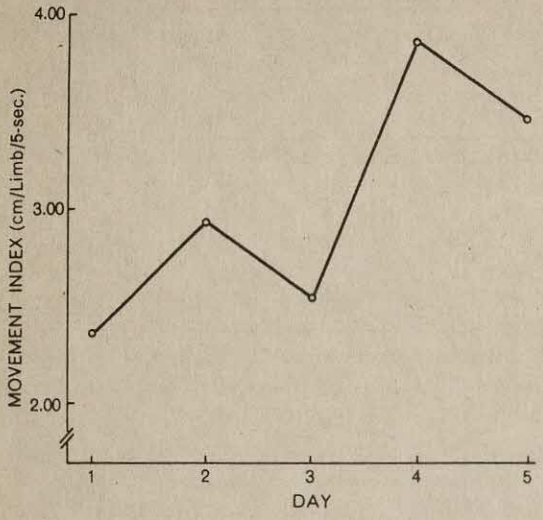


Figure 2 shows changes over days in mouthing, hand-mouth contacting, and crying. As might be anticipated from the earlier finding of great variability of mouthing and crying, neither of these measures shows significant day effects. Whether the increase in crying scores between day 1 and day 4 which appears to visual inspection can be brought to significance by a larger sample, whether of subjects of time, remains to be investigated. Of more interest is the finding that mouthing is insignificantly different from day 1 to later days. It might have been anticipated that feeding experience would affect mouthing, either increasing it as a function of the reinforcement operation of feeding or decreasing it as a function of the increasing frustration of nonnutritive mouthing. Neither effect is evident.

The hand-mouth contacting measure is somewhat more difficult to interpret. The differences, shown in Figure 2 are significant ($F = 4.51$, $df = 14/56$, $p < .01$), although it is apparent that the sharp increase in contacting on day 3 accounts for most of the variation. An examination of other responses and of hospital routines has shown no covariation which would readily account for this surprising hump. Nonetheless, it is interesting that, as in the case of mouthing, the increasing number of feeding "trials" the child has does not seem to influence hand-mouth contacting in any systematic way. Contacts on

Fig. 3: *Changes in neonatal movements over the first five days of life.*



day 1 are as frequent as those on day 2, day 4, or day 5. This finding is important in that it casts serious doubt on an assumption that hand-mouth contacting is "secondary" to feeding or learned in the feeding situation.

The day-to-day plot of movement for group IIa subjects shown in Figure 3 is relatively free of ambiguity. There is a systematic and significant ($F = 3.03$, $df = 4/36$, $p < .05$) increase in neonatal movement during the early days of postnatal life. This finding, beyond confirming the observations of other investigators, adds to our confidence that the movement measure used in the present research is a sound measure of individual and situational variation.

Interrelations among measures. Do the response measures hang together in a way which would suggest a primitive "typology" of newborns' For each child in group II ($N = 15$), a single index was computed of crying, mouthing, contacting, and movement by summing daily values over the entire hospital stay. The rank correlations (rho) among these *over-all* measures were as follows:

Contacting-Mouthing	.36
Contacting-Crying	.19
Contacting-Movement	.19
Mouthing-Movement	.49
Movement-Crying	.18
Mouthing-Crying	−.27

Although none of these correlations reaches conventional levels of statistical significance, there is a suggestion that the more active infants mouth more and that the infants who mouth more often make contact more often. There is, furthermore, no strong evidence that a simple "general factor" will account for the relations suggested. It is interesting to note that the low correlation between movement and contacting argues against a random-hit interpretation of hand-mouth exchange; it does not seem to be the case that an infant who moves a lot happens to strike his mouth frequently.

The complexity of interrelations among newborn responses is seen in an examination of day-to-day changes in correlations (Table 1). For the comparisons (again, the coefficient is rho) crying was not included because of the relatively high incidence of zero scores. The major implication of this table is clear; a tendency for the three measures to be related fairly closely early in the postnatal period dissipates toward chance by the fifth day of life. Two hypotheses can be advanced to account for this striking shift. The interrelation at birth and shortly thereafter may be under the control of some transiently effective variable (e.g., maternal anesthetic), or the environmental contingencies of the first five days may be such as to disturb a congenitally organized response-set. The decay in interrelations among measures cannot be assigned to increasing unreliability of measures. For both movement and frequency of contacting, the *individual* stabilities between day 4 and day 5 are as great as or greater than the stability among earlier days. An accurate interpretation of these findings must await further empirical examination of newborn behavior and its environment.

The Effects of Stimulation

On two observations each day, the infants in group IIa were given an opportunity to suck on the observer's shielded index finger for 10 to 15 seconds. Under the conditions of the present study, no stable immediate effect on neonatal movement could be detected; although the infants seemed to quiet somewhat when the stimulation was present, this general observation was not confirmed by an examination of the film-strips.

Sex differences. Some variation between the behavior of males and the behavior of females

Table 1: *Rank-order correlation coefficients (RHO) from day to day among three measures of newborn behavior*

	Movement-Contacting	Movement-Mouthing	Contacting-Mouthing
Day 1	.57*	.55*	.68**
Day 2	.25	.35	.59*
Day 3	−.09	−.21	.70**
Day 4	−.16	.17	.32
Day 5	.11	−.03	.18

Note.—One asterisk indicates significance at .05 level, two asterisks significance at .01 level. N = 16 for days 1, 2, 3, and 4; N = 13 for day 5.

was noted in each group but no sex differences were found to be both statistically significant and replicated.

DISCUSSION

Two of the measures of newborn behavior under study, movement and hand-mouth contacting, met reasonable demands for reliability of observation and for sensitivity to individual variation. The other measures, mouthing and crying, can be observed in the neonate with satisfactory precision, but did not show in the setting described here significant sensitivity to individual differences. Of the four, movement showed systematic changes over the lying-in period.

The finding of stable individual differences in movement and hand-mouth contacting among human newborns confirms the often stated but infrequently demonstrated proposition that human beings show substantial variations one from another aboriginally. Even for the relatively homogeneous group of children observed for this research, wide and stable differences were shown. Knowing this, it is important to turn to two further questions. Do these early differences make any difference later? Does it make good sense to consider these variations, whether from time to time or from baby to baby, as representing differences in tensions? The first of these questions requires longitudinal study; the second demands an examination in some detail of the relation between activity and contacting on one hand and environmental events and other neonatal responses on the other. If neonatal movement is to serve as an index or partial interpretation of the term "tension," then it should be possible to demonstrate systematic relations to privation, to noxious external stimulation, and to the environmental operations used by parents and clinical practitioners to reduce distress in infants. On the basis of the present study, we can proceed to an attack on these next empirical questions with some confidence in the precision and the stability of measures for hand-mouth contacting and infant movement.

SUMMARY

An initial attack on the problems of newborn behavior was made in an examination of the reliability, individual stability, and relatedness of four response measures, movement, crying, mouthing, and hand-mouth contacting. Infants were seen repeatedly during the lying-in period. All measures showed very high intersubjective reliability. The measures for movement and hand-mouth contacting were sensitive to individual variation over the first five days of life; the measures for crying and mouthing were not. The movement measure showed systematic change over the lying-in period; for the other measures, neither a regular increase nor a regular decrease in occurrence was found, putting in some doubt the suggestion that mouthing and hand-mouth contacting are phenomena secondary to feeding.

The relation of the findings to the problem of tension and tension-control in the newborn was discussed.

REFERENCES

DENNIS, W. A description and classification of the responses of the newborn infant. *Psychol. Bull.*, 1934, *31*, 5–22.

FREUD, S. *Three essays on the theory of sexuality.* London: Allen & Unwin, 1905 [1949].

GARDNER, G. E. Problems of early infancy. *J. Amer. Psychoanal. Ass.*, 1955, *3*, 506–514.

GESELL, A. *The embryology of behavior.* New York: Harper, 1945.

IRWIN, O. C. The amount and nature of activities of new-born infants under constant external stimulating conditions during the first ten days of life. *Genet. Psychol. Monogr.,* 1930, *8,* 1–92.

KESSEN, W., HENDRY, A. L. S., & LEUTZENDORFF, A. M. Measurement of movement in the human newborn: a new technique. *Child Develpm.,* 1961, *32,* 95–105.

LUSTMAN, S. L. Rudiments of the ego. *Psychoanal. Study Child,* 1956, *11,* 89–98.

PIAGET, J. *The origins of intelligence in children.* New York: International Universities Press, 1952.

SPITZ, R. A., & WOLF, K. M. Autoerotism. *Psychoanal. Study Child,* 1949, *3/4,* 85–120.

WEISS, A. P. The measurement of infant behavior. *Psychol. Rev.,* 1929, *36,* 453–471.

WOLF, K. M. *The origins of individuality.* Mimeo., 1957.

12

The Developmental Behavioral Approach to the Neurologic Examination in Infancy

Hilda Knobloch and Benjamin Pasamanick

Knobloch and Pasamanick are physicians who have engaged in considerable research on the cause and effects of organ deficiencies (particularly brain deficiencies) in the human child. Their viewpoint in this paper is more maturational than is usual in the current psychological articles. They stress the need for the physician to know the processes of human development and for the infant to participate actively in the examination.

In what may be a reaction against the so-called "dynamic" approach in neuropsychiatry, it has become almost a fetish to talk about "brain damage." To some this means defeatism; to others, ourselves included, the outlook in the response to treatment for most children with organic brain disease is sometimes more hopeful than in those whose disorders result from gross social pathology. I need not belabor, then, the importance of early diagnosis from the point of view of treatment, prevention, and research.

Brain damage may manifest itself in a variety of neuropsychiatric abnormalities, and the term does not mean the same thing to all people. Within the medical specialties concerned with this problem, the confusion is as great as it is among other professional workers. Nevertheless, we believe that the physician has the responsibility for the specification of diagnostic criteria and for the development of additional diagnostic tools.

A good diagnostic tool must fulfill three important demands. First, it must be reliable, that is, give the same results when used by different people. Secondly, it must be valid, that is, able to predict later development. Thirdly, it must be able to tap the patterns of behavior in which abnormality occurs. We plan to present data bearing on these three points in relation to the Gesell Developmental Examination (1954), but first we would like to present some necessarily compressed theoretical background. This is not a finished structure, and not all the pieces are fastened in place, but we believe it does provide a firm basis for going forward.

There is a considerable body of evidence concerning the functional development of the central nervous system which indicates that the most opportune time for detecting deviations from normal neuropsychologic development is the period of infancy and that diagnosis becomes increasingly difficult with advancing age. The evidence also demands reconsideration of the effectiveness of the classical neurologic approach to the problems of differential diagnosis and the need for developing additional diagnostic instruments.

Systematic observations of large numbers of normal and abnormal infants have demonstrated that development proceeds in a predictable and orderly fashion. The form that behavior takes as the organism attains maturity can be delineated quite precisely in terms of direction, tempo, interrelationships, and complexity.

One of the most important facts to be borne in mind in the evaluation of neurologic status and the differentiation of retardation from distortion is the intimate interrelationship of neurologic integrity and maturational level. At all times, the chronological age (or the corrected chronological age in premature infants) must be considered in evaluating the significance of the

---

From *Child Development,* 1962, *33,* 181–198. Reprinted with permission of the Society for Research in Child Development.

behavior patterns observed. At any one stage of maturity, the patterning to be expected at a later stage can be predicted from the behavior observed at this earlier age level. We can differentiate between normal, retarded, and distorted function. This task is made easier because brain damage is usually diffuse and pervades the whole central nervous system. By way of illustration, at 16 weeks of age the legs are predominantly passive and no abnormal neurologic patterning in the lower extremities may be evident, even though disease is present. Yet one knows that, if neurologic damage exists, abnormal patterns in organization and control will be present at 40 weeks when the legs are active participants. The nature and extent of these abnormalities can be predicted from the type of head, arm, and upper trunk control achieved at 16 weeks of age.

Another major factor in the evaluation of the significance of abnormal behavior is the fact that compensation for neuromotor disability occurs in the process of maturation itself. There is, if you will, "a drive towards normality." The nature of the process by which this occurs is purely speculative. The greatest degree of involvement usually occurs in those areas in which the most active organization and individuation is taking place and which are coming under voluntary control at that particular developmental stage. For example, at 16 weeks of age abnormalities are seen primarily in head and eye-hand control. By 28 weeks a modicum of head control may be achieved while there is still great difficulty in integrating arm and upper trunk movements. By 40 weeks of age head control usually is, and arm control may be, normal, but precision in the use of the fingers still lags, independent sitting is not attained, and there is little integration of the lower extremities into the total behavior picture. Perhaps the chemical and electrical processes concerned are so active that abnormal excitation-inhibition relationships are present which have a disorganizing effect, and only after the vortex of activity has moved on to another area can the integration that is normally expected take place. The maturational process is such, however, that the manner and the rate at which compensation will occur can also be predicted from the nature of the abnormalities observed at any given stage. If hypotonicity is the major manifestation of the disease process, a child who is not sitting or taking weight on his feet at 40 weeks of age will usually be walking by 24 months. An athetoid who supports his body weight at 9 to 12 months, and who no longer has wild plunging of his head, will walk sooner than a child of the same intellectual ability who still has forward sagging of his head and who slides through the examiner's arms and collapses his legs when he is placed in standing. These illustrations are of marked deviations, but the same principles apply to lesser degrees of neurologic abnormality.

In infancy the classic neurologic examination, because of the limited stimulus-response approach, fails to tap the areas of behavior that we wish to evaluate. For minimal degrees of damage this is true in older children as well. This examination fails to elicit optimal responses, or often any responses at all, primarily, but not exclusively, because the infant is too immature to comprehend or to execute what is expected. This process of applying a stimulus and eliciting a specific response is inadequate because of its very rigidity. By and large, it permits only one response, which is either present or absent, and often this presence or absence can not be interpreted.

In contrast, the behavioral examination situations and test objects are so designed that characteristic responses are expected at each developmental stage in each situation. This approach permits differentiation and detailed exploration of abnormal as well as normal responses. As Arnold Gesell has said, functional tests of behavior define the maturity and the integrity of the central nervous system.

The inadequacy of the classical neurologic examination is readily apparent in the newborn period, as evidenced by the failure of this examination to diagnose hydranencephaly. This failure has given rise to a great deal of expressed concern over the problem of "evaluating the function of an organ which does not function" (Masland, 1960). It is based on the fallacy that no cortical activity exists at birth. There is, in fact, evidence in serial electroencephalograms of premature infants that, at given stages, patterned electrical activity becomes apparent whose subsequent development, as well as the emerging form, can be delineated. Granted that the behavioral repertoire of the newborn infant is very limited, there are two items of behavior which not only are clearly present by the end of the first week but also progress in a very specific and predictable fashion throughout the developmental period. Visual fixation and auditory attentiveness are not present in hydranencephalic infants at any age. They are present in normal newborn infants, and their extent and variability from infant to infant can be delineated. Furthermore, the increasing complexity of patterns that will emerge at 6 weeks, 8 weeks, 12 weeks, and beyond can be defined. In addition, the fluidity of the behavior of the normal newborn infant is replaced in the hydranencephalic by stereotyped and repetitive reflex responses which are qualitatively different. Cortical activity does not arise "de novo" at some indefinite point in the postnatal period, which those who are proponents of its absence at birth are unwilling to specify. Yet neither visual nor auditory attentiveness is included in the neurologic examination of the neonate by those currently writing about it or producing movies of the examination technique.

A discussion of the importance of the developmental approach is incomplete without consideration of the role that this concept plays in the examination procedure itself. All too often the examiner thinks that the infant can be nothing more than a passive participant, and the baby is placed on his back and his extremities are manipulated. This procedure is partly a result of the structure of the standard neurologic examination, but it must also stem from the physician's realization that the infant has a limited understanding of language. In order to evaluate the infant's behavior adequately, his active participation in the examination is essential. Because of the fact that the newly emerging patterns are accompanied by supporting accessory patterns which facilitate the expression of the new behavior, an infant with a handicap may appear normal if he is merely placed in supine and manipulated. The abnormality may become apparent only if the infant is stimulated to reach out for something or if he is placed in sitting or standing as well as in supine and prone. We must attempt to elicit all of the behavior expected in a normal infant of the same age. The baby must, so to speak, be placed under "stress" and his active participation invoked. This is best done by giving him toys, which are in fact specific test objects, and watching what he does.

The age at which the infant is examined may greatly modify the behavior that is observed. Prior to 16 weeks of age, for example, the reflex quality of neonatal behavior has not entirely disappeared, and an infant with neurologic defect may have good head control in the prone position because of the predominance of extensor tone. If he is not placed in sitting and expected to look down at a table top, we may fail to realize that his head is retracting. As such an infant grows older and the reflex extensor tone disappears, the spuriously good head control also disappears. The infant is no longer able to lift his head from the bed in the prone position, and when he is placed in sitting his head plunges forward dangerously and he has great difficulty in re-erecting it. Even reflex behavior, then, is modified by the maturational process.

Time does not permit a detailed analysis of the emergence of a specific behavior pattern to show the interrelationships with other patterns and the importance of accessory supporting behavior in the developmental process. It is necessary to re-emphasize at this point that we cannot merely make a list of abnormal signs and evolve a diagnosis. We must relate them not only to the level of maturity of the infant but also to all of the other behavior patterns that the infant manifests. A thorough knowledge of normal growth and development is the basis on which detection of abnormality rests.

In this connection at least, passing mention must be given to another essential factor, namely the importance of the monistic approach. We cannot fragment the infant into psychometric and neurologic halves. In this as in other areas of neuropsychologic functioning, it is the responsibility of the physician, with his knowledge of disease processes, to make a diagnosis, even if it is one of "no disease." It is, we believe, incorrect to state that the physician needs only a superficial acquaintance with development; that he does the neurologic examination, but if he wants a really competent evaluation of the maturity of the child, let us say because adoption is pending, the infant should be sent to a good psychometrist.

By and large, neurologic abnormality in infancy is diffuse rather than localized, and there is a continuum of damage which ranges from minimal to severe. Damage to the brain can give rise to a multitude of abnormalities, all of which must be carefully evaluated when searching for evidence of injury. All have maturational aspects and these undoubtedly have not been completely explored; we can do no more than mention some: seizure states, particularly the abortive seizures; emotional stability; attentional patterns; organization, integration, and autism; auditory, receptive, and interpretive integrity; visual function; sensory perception; reflex responses. These are other factors basic to adequate evaluation.

We would like to talk about two of the major manifestations of organic brain disease—mental deficiency and the multitude of conditions called cerebral palsy—to clarify definitions, delineate the types of abnormalities found, since it is es-

Table 1: *Correlations between examiners for clinical cases examined by Gesell Developmental Examination*

	EXAMINERS							
	1 & 2		1 & 3		1 & 4		3 & 4	
	N	*r*	*N*	*r*	*N*	*r*	*N*	*r*
General Developmental Quotient	20	.995	44	.995	37	.98	25	.98
Neurologic Status	14	.995	40	.95	29	.94	18	.99
Intellectual Potential	15	.98	40	.99	28	.98	15	.98

sential to distinguish between retardation and abnormal deviation, and demonstrate some aspects of the continuum of damage. The most severe deviation in the continuum, "cerebral palsy," is usually considered as a qualitative change independent of the intellectual potential of the child, although of course highly associated with mental deficiency. It involves reflex responses and muscle tonus and control in postural adjustments, locomotion, and manipulation. For the sake of brevity, we will use the terms motor or neurologic to include all of these aspects in the discussion which follows. Careful examination of behavior discloses that the same abnormalities are found in the lesser degrees of neurologic involvement as in cerebral palsy and that there is a point where quantitative changes become qualitative. This is probably the point at which there is a difference in the prognosis for later functioning, and in the response to a classical neurologic examination to be expected at older ages. The evaluation of motor neurologic status in infancy carries with it a prognosis which is itself important in delineating the diagnostic categories. When we make a diagnosis of cerebral palsy in infancy, we mean that the standard neurologic examination at school age will show definite abnormality, even though it may be so mild no treatment is required. In the remainder of the cases in the continuum, it is necessary to distinguish between those children who have motor retardation, that is, a gross motor developmental level significantly below their chronologic age because of generalized retardation, and those who have abnormal neurologic integration in the presence of an intellectual potential within the normal range. There are parallel degrees of neurologic deviation in the two groups, but the prognosis for those with mental deficiency is of course poorer. It is particularly in this latter group that we have to bear in mind the functional level of maturity in evaluating the neurologic findings.

Four degrees of neurologic integrity should be considered in the rest of the continuum: (a) Abnormal neurologic signs of marked degree. In this category motor behavior is considerably retarded and abnormal, and, if the findings were present in an older child, a diagnosis of cerebral palsy would be made. Clinical experience indicates, however, that the majority of such infants compensate, and a standard neurologic examination at school age shows no neurologic residua. Abnormalities *are* evident if the development is compared with that appropriate to chronologic age. (b) Abnormal neurologic findings of minor degree—the "minimal cerebral damage" category. These infants show the characteristic deviations but to a less marked degree, with gross compensation by 18 to 24 months for the standard neurologic examination, but again persistence of deviations can be demonstrated by a developmental evaluation. (c) Abnormal neurologic signs of no clinical significance. Some abnormal neurologic patterns are present but it is questionable if they are sufficiently marked to place the child significantly outside the range of normal. (d) Normal neurologic functioning.

You have finally, in this last section, heard the term minimal cerebral damage, the presumed topic of this symposium. We have purposely gone into this much detail in order to present minimal injury in the context of the broader framework of which it is such an integral part. We have specified our criteria for separating minimal from marked abnormality, and they are available (1959). If we have not made it clear, however, that minimal damage is only one part of a continuum in which the diagnosis that is made, on the basis of the observed manifestations of damage, varies with the age and maturity level of the child, then we have failed hopelessly in what we have set out to do.

There is one additional point we must make before proceeding to give you the data which have been accumulated in support of the broad concepts just discussed. The examination about which we have been talking is the Gesell Devel-

Table 2: *Correlations between examiners for longitudinal study cases examined by Gesell Developmental Examination*

	EXAMINERS			
	1 & 3		1 & 4	
	N	r	N	r
General Developmental Quotient	12	.96	23	.96
Neurologic Status	10	.95	22	.88
Intellectual Potential	10	.90	19	.95

opmental Examination. It is primarily a neurologic tool; without it only the grossest deviations from normal will be detected. Relatively speaking, the behavior repertoire throughout infancy is restricted, and there is a limit to how many things are available for incorporation in any examination. The fact that a number of the people who are currently writing on pediatric neurology have not seen fit to give credit to Dr. Gesell is distressing. Perhaps it is because they are not using correctly the parts of his examination that they have appropriated.

We have already presented data on the validity of the Gesell Developmental Examination for predicting later *intellectual* functioning on several occasions (1960). Correlations between examinations in infancy and those at 3, 5, and 7 years range from .5 to .75. . . . The more abnormal the infant, the higher the correlations, a very desirable characteristic in a clinical tool.

Data on the reliability of the examination have been obtained in two ways. In the first method, one examiner does the examination, which is observed from behind a one-way vision screen by the other physicians, and the diagnoses are made independently. Table 1 indicates that, for a series of clinical cases referred for evaluation, correlations between four different examiners for neurologic status, intellectual potential, and general developmental quotient range from .94 to .995. For a series of essentially normal cases seen in a longitudinal study, they are slightly lower, ranging from .88 to .96. These data are presented in Table 2.

In the second approach to reliability, the same infant is examined successively by two examiners independently. The infants in this group were also essentially normal babies ranging in age from 9 to 12 months. They spent at least one hour in the examination situations. This method of evaluation takes into account not only interexaminer differences but also variations in the behavior of the infant. The stability of infant behavior is itself important. Interexaminer correlation in this series is .83 for neurologic status and .82 for general developmental quotient. All of these correlations are highly significant.

The other major question to be answered about the neurologic diagnostic tool is whether or not it is capable of eliciting responses in those areas of behavior that are abnormal. Are there items in other areas of behavior to which the motor neurologic diagnosis is related? Does the neurologic diagnosis at 40 weeks predict the development of abnormal behavior in later childhood? Is there a relationship between the qualitative diagnosis of neurologic status and the quantitative evaluation of the amount of abnormality observed?

A detailed study (Knobloch & Pasamanick, 1959) was made of 46 individual neurologic patterns in order to see the relation between the amount of abnormality recorded and the clinical neurologic diagnosis made. The neurologic patterns were rated as 0, or normal, and 1, 2, and 3+, representing various degrees of abnormality. For each of the neurologic categories the average score of all of the cases for all 46 patterns was obtained. The results shown in Table 3 indicate that the mean score increases significantly as the severity of the estimate of neurologic status increases. The mean score for the normal group is 3.9, for the possible minimal damage group (abnormal neurologic signs of no clinical significance), 11.1; these two groups are combined as the normal category in later tables. For the minimal damage group (abnormal neurologic signs of minor degree) it is 20.9. For the possible cerebral palsy group (abnormal neurologic signs of marked degree) it is 24.8, and for the overt abnormality group (cerebral palsy), 27.7; these two groups comprise the markedly abnormal category.

. . . We would like to re-emphasize that the abnormalities on which the diagnosis of minimal damage is based are the same ones found in cerebral palsy. For example, difficulty in retaining, fisted hands with adduction of the thumb into the palm (the so-called "cortical thumb"), mal-

Table 3: *Mean score on individual neurologic patterns according to neurologic diagnosis at 40 week examination*

Neurologic Status	*N*	*Mean*	*SD*
Normal	680	3.9	4.4
Possible Minimal Damage	121	11.1	5.8
Minimal Damage	137	20.9	7.5
Possible Cerebral Palsy	37	24.8	8.4
Overt Abnormality	17	27.7	11.5

Table 4: *Incidence of abnormal behavior during first month related to neurologic status at 40 weeks*

	NEUROLOGIC STATUS				
Abnormal Behavior	*Normal* *N* = 801	*Minor Abnorm.* *N* = 137	*Marked Abnorm.* *N* = 54	*Total* *N* = 992	*p*
	Percentage of infants with abnormal behavior				
Sucking Difficulty	11.0	16.9	27.8	12.7	< .01
Hypotonicity	4.7	9.5	14.8	6.0	< .01
Hypertonicity	7.3	10.3	18.5	8.3	< .05
Hypo- or Hyper-tonicity	12.1	19.8	33.3	14.3	< .01
Vomiting	13.3	18.4	20.4	14.4	< .20
Crying Excessively	14.3	11.0	16.7	14.1	< .70

directed reaching, and sitting on a narrow base with adduction and extension of the legs are seen in almost all infants who are diagnosed clinically as having cerebral palsy. Scientific parsimony would dictate that, when these same abnormalities are seen in those infants who are diagnosed as having minimal damage, they should be ascribed to the same cause to which they are attributed in cerebral palsy, particularly when they are significantly associated with the same etiologic factors, such as prematurity. It seems most unreasonable to say that in the case of cerebral palsy they are due to brain damage, but to invoke some other and unspecified explanation for their presence when they are found in those infants with insufficiently severe or numerous patterns to permit a diagnosis of cerebral palsy. The relation between the neurologic findings and other aspects of the behavior at 40 weeks and at 3 years also form part of an integrated structure with the brain as the focal point. We rarely have complete evidence for validation of any hypothesis. For example, although we have neither pathologic findings nor demonstration of the specific virus, for operational purposes there is no difficulty in accepting the concept that roseola is a virus disease.

At the time of the 40 week examination historical information was gathered, independently from the physician's evaluation, about various behavior patterns in the first month of life and in the time up to the examination. Analysis of these data indicates that, in general, as the neurologic abnormality increases, the incidence of an abnormal pattern also increases. As Table 4 demonstrates, in the first month there are significant differences between the normal, minor abnormality, and marked abnormality groups in the incidence of sucking difficulties, hypotonicity, and hypertonicity. Table 5 indicates that up to the time of the 40 week examination there are significant differences in the incidence of twitching and the number of feeding problems and in the percentage of infants described as usually cranky.

When the mother brought the infant for examination at 40 weeks of age, an independent interviewer made a qualitative estimate of whether or not the mother was relaxed, self-confident, tense, or uncertain. This qualitative estimate was correlated with the abnormal behavior manifested by the infant as obtained by history, as well as with the neurological diagnosis made by the physician. Table 6 summarizes the associations of abnormal neurologic status and tension in the mother with other abnormal behavior in the in-

Table 5: *Incidence of abnormal behavior up to time of examination related to neurologic status at 40 weeks*

	NEUROLOGIC STATUS				
Abnormal Behavior	*Normal* $N = 801$	*Minor Abnorm.* $N = 137$	*Marked Abnorm.* $N = 54$	*Total* $N = 992$	*p*
	Percentage of infants with abnormal behavior				
Feeding Problem	6.4	11.7	18.5	7.9	$< .01$
Crankiness	1.7	3.7	9.3	2.4	$< .01$
Convulsions	1.9	3.6	5.6	2.3	*
Twitching	4.1	5.2	13.0	4.7	$< .01$
Fear of Strange Things	3.5	2.2	7.4	3.5	$< .30$
Sleeping Problem	5.5	5.8	5.6	5.6	$< .90$
Startling	12.3	19.0	25.9	15.9	$< .20$
Difficulty with Solids	4.0	1.4	7.4	3.8	$< .30$

** Differences between normal and marked abnormal groups significant at .05 level.*

Table 6: *Associations of abnormal neurologic status and tension in the mother with abnormal behavior in the infant*

Infant Behavior	*Abnormal Neurologic Status*	*Mother Tense*
Sucking Difficulty	+	+
Hypotonicity	+	−
Hypertonicity	+	+
Hypo- or Hypertonicity	+	+
Vomiting	−	+
Crying Excessively	−	−
Feeding Problems	+	+
Crankiness	+	+
Convulsions	+	+
Twitching	+	+
Fear of Strange Things	−	−
Sleeping Problem	−	+
Startling	−	+
Difficulty with Solids	−	−

Note.—"+" indicates statistically significant differences.

fant. Significant associations over the range of abnormal neurologic status or tenseness are found in all except three of the items: crying excessively, fear of strange things, and difficulty in taking solids. A trend is present, however, and the markedly abnormal group is different from the other two.

More items are associated with tenseness than with the neurologic diagnosis, but these two factors are not independent, as inspection of Table 7 reveals. The percentage of mothers called tense increases over the five neurologic diagnostic categories from 2.8 per cent in the normal group through 4.1, 5.8, 10.8 to 29.4 per cent in the overtly abnormal group. An additional item included in this table is the relation between neurologic diagnosis and illness in the infant. Over the five diagnostic categories the percentage of infants who had illness increases from 22.5 through 39.6, 41.6, 56.7, to 64.5. This is undoubtedly an additional factor contributing to the highly significant association between the emotional state of the mother and the neurologic diagnosis made on her infant. These data raise some subsidiary problems for investigation. Since behavior problems in children are very often attributed to tension in their mothers, serious consideration should be given to the other side of the coin, namely, whether or not the infant with difficulties produces the tension in his mother.

A final aspect on which data are available concerns the relation between neurologic status at 40 weeks and certain aspects of behavior at 3 years. Of the original 1000 infants seen at 40 weeks of age, approximately 300 were re-exam-

Table 7: *Incidence of illness and percentage of mothers called tense, related to neurologic status at 40 weeks*

Neurologic Status	*N*	*Mother Tense*	*Illness in Infant*
		%	%
Normal	680	2.8	22.5
Possible Minimal Damage	121	4.1	39.6
Minimal Damage	137	5.8	41.6
Possible Cerebral Palsy	37	10.8	56.7
Overt Abnormality	17	29.4	64.5
Total	992	4.1	32.6
p		$< .01$	$< .01$

Table 8: *Relation between neurologic status at 40 weeks and change in intellectual potential from 40 weeks to 3 years (All Infants)*

	Neurologic Status at 40 Weeks			
Intellectual Potential	*Normal*	*Minor Abnorm.*	*Marked Abnorm.*	*Total*
No. Infants Low Average or Better at 40 Weeks	224	36	9	269
% Remaining Low Average or Better at 3 Years	93.6	86.1	66.7	91.8
% Falling to Below Average at 3 Years	6.4	13.9	33.3	8.2

Note.—$\chi^2 = 6.55$; $p < .02$ with 1 df (minor and marked abnormality groups pooled).

ined by the same examiner at 3 years of age, almost wholly without knowledge of the 40 week status. With a small number of abnormal cases in a sample, statistically significant differences are difficult to demonstrate. The fact that, in spite of this, we can show a large number of differences significant at the .01 level merely attests to the validity of the instrument.

As would be expected, a smaller proportion of the children with marked neurologic abnormality had low average or better intellectual potential at 40 weeks of age. Table 8 indicates the more important significant relation between the neurologic status at 40 weeks and the percentage of children remaining low average or better in intellectual potential at 3 years. If the 40 week neurologic status was normal, 94 per cent of the children remained low average or better at 3 years. The corresponding figures for those with minor abnormality and marked abnormality at 40 weeks are 86 per cent and 67 per cent. We might postulate that the examination situation constitutes a stress for a 3-year-old child, and those infants with disordered cerebral function are less able to cope with the situation. Table 9 indicates that the differences over the three neurologic diagnostic categories are significant for the nonwhite infants, ranging from 93 per cent through 71 per cent down to 67 per cent. The differences for the white infants are not statistically significant, and we can postulate a number of reasons why the examination situation might be more stressful for the Negro children, e.g., the presence of a white examiner and a difference in language amounting almost to a dialect being only two. Table 10 shows that the differences are also significant for the premature infants, the percentages for the neurologic diagnostic categories ranging from 95 through 86 to 67. The fact that these differences are not significant in the control group is probably related to two factors other than the prematurity itself. A much smaller percentage of the control infants had abnormal neurologic status at 40 weeks of age; only three, for instance, had marked neurologic abnormality. Furthermore, a much smaller percentage of the control infants who were re-examined were white. It is important to point

Table 9: *Relation between neurologic status at 40 weeks and change in intellectual potential from 40 weeks to 3 years (Nonwhite Infants)*

	Neurologic Status at 40 Weeks			
Intellectual Potential	*Normal*	*Minor Abnorm.*	*Marked Abnorm.*	*Total*
No. Infants Low Average or Better at 40 Weeks	123	14	6	143
% Remaining Low Average or Better at 3 Years	92.7	71.4	66.7	89.5
% Falling to Below Average at 3 Years	7.3	28.6	33.3	10.5

Note.—$\chi^2 = 9.41$; $p < .01$ with 1 df (minor and marked abnormality groups pooled).

Table 10: *Relation between neurologic status at 40 weeks and change in intellectual potential from 40 weeks to 3 years (Premature Infants)*

	Neurologic Status at 40 Weeks			
Intellectual Potential	*Normal*	*Minor Abnorm.*	*Marked Abnorm.*	*Total*
No. Infants Low Average or Better at 40 Weeks	83	14	9	106
% Remaining Low Average or Better at 3 Years	95.2	85.7	66.7	91.5
% Falling to Below Average at 3 Years	4.8	14.3	33.3	9.5

Note.—$\chi^2 = 6.34$; $p < .02$ with 1 df (minor and marked abnormality groups pooled).

out at this juncture that the 300 cases re-examined did not represent any selection but were merely the first infants in the series.

Another aspect of behavior at 3 years of age that shows a relation to the neurologic diagnosis at 40 weeks is the achievement of toilet control. Table 11 shows that the more normal the neurologic function the better the control. The percentage of infants achieving bowel control at 3 years of age is 89 per cent if the 40 week neurologic status was normal, 79 per cent if there was a minor abnormality, and 64 per cent if there was a marked abnormality at 40 weeks. The corresponding figures for daytime urinary control are 78 per cent, 74 per cent, and 41 per cent, and for nighttime urinary control, 60 per cent, 42 per cent, and 45 per cent. The same trend over the five 40 week neurologic diagnostic categories is evident in the median age at which toilet control was achieved and is shown in Table 12. For bowel control the median age for the normal group is 18 months, rising to 24 months for the overt abnormality group. For daytime urinary control the figures are 18 months and 27 months and for nighttime urinary control, 20 months and 26 months. The small number of cases involved undoubtedly accounts for the fluctuation observed. It is probably also the reason why the differences are significant only because the markedly abnormal group does so poorly, although the trend for the continuum is evident.

At 3 years the physician again made qualitative judgments about the behavior of the child during the examination itself, and the proportion of children showing abnormal behavior was correlated with the 40 week neurologic diagnosis. Table 13 shows those items in which there is a trend over the three 40 week diagnostic categories. These are almost all integrative characteristics—organization, interest, maturity, discrimination, judgment, exploitation, spontaneity, and the over-all quality of the integration of behavior. That these are areas of behavior which are abnormal in the "brain-damaged" child is obviously not coincidence. The differences are statistically significant except for interest and spontaneity. In Table 14 are those items in which there is either a dichotomy or no difference between the children markedly abnormal at 40 weeks and the remainder. These are primarily emotional characteristics—stability, restlessness, crying, dependency on the mother, and an overall evaluation of emotional stability. Also included are attention, hyperactivity, and scatter. Differ-

Table 11: *Relation between neurologic status at 40 weeks and percentage of 3-year-olds achieving toilet control*

	Neurologic Status at 40 Weeks			
Achievement of Control at 3 Years	*Normal*	*Minor Abnorm.*	*Marked Abnorm.*	*Total*
Bowel*				
No. with Neurologic Diag.	235	38	22	295
% Achieving Control	88.8	79.0	63.6	85.8
Daytime Urinary†				
No. with Neurologic Diag.	233	38	22	293
% Achieving Control	78.2	73.7	42.1	74.8
Nighttime Urinary‡				
No. with Neurologic Diag.	233	38	22	293
% Achieving Control	63.6	41.0	45.4	56.6

* $\chi^2 = 13.95$; $p < .01$ with 2 df.
† $\chi^2 = 78.61$; $p < .01$ with 1 df (minor and marked abnormality groups pooled).
‡ $\chi^2 = 6.24$; $p < .05$ with 2 df.

Table 12: *Relation between neurologic status at 40 weeks and median age of achieving toilet control*

	ACHIEVEMENT OF CONTROL					
	Bowel		Daytime Urinary		Nighttime Urinary	
Neurologic Status at 40 Weeks	*No. Achiev-ing*	*Median Age* (Mos.)	*No. Achiev-ing*	*Median Age* (Mos.)	*No. Achiev-ing*	*Median Age* (Mos.)
Normal	186	17.7	176	18.0	150	19.7
Possible Minimal Damage	24	23.5	22	18.0	21	19.0
Minimal Damage	37	17.5	33	22.4	25	22.3
Possible Cerebral Palsy	10	22.0	7	24.0	6	18.0
Overt Abnormality	4	24.0	4	27.0	5	26.0
Total	251	17.7	242	19.4	207	19.9

Table 13: *Relation between neurologic status at 40 weeks and behavior at 3 year examination: items with trend over all diagnostic categories*

		NEUROLOGIC STATUS AT 40 WEEKS			
Behavior at 3 Years	*N*	*Normal*	*Minor Abnorm.*	*Marked Abnorm.*	*p*
		Percentage of infants with abnormal behavior at 3 years			
Organization	288	16.7	18.9	65.2	< .01
Interest	290	14.4	15.8	30.4	*ns*
Maturity	282	22.3	31.5	65.2	< .01
Discrimination	282	16.5	28.6	52.2	< .01
Judgment	284	10.7	16,2	52.2	< .01
Exploitation	283	9.8	16.7	43.4	< .01
Spontaneity	291	20.0	21.0	26.1	*ns*
Quality of Integration	286	15.1	18.9	60.8	< .01

Table 14: *Relation between neurologic status at 40 weeks and behavior at 3 year examination: items with differences between marked abnormality and rest or with no differences*

		NEUROLOGIC STATUS AT 40 WEEKS			
Behavior at 3 Years	*N*	*Normal*	*Minor Abnorm.*	*Marked Abnorm.*	*p*
		Percentage of infants with abnormal behavior at 3 years			
Stability	288	14.0	8.3	34.7	< .01
Attention	288	27.6	24.3	78.2	< .01
Perseveration	291	8.2	5.4	39.0	< .01
Irritability	293	3.1	2.6	21.7	< .01
Restlessness	292	18.1	10.5	47.8	< .01
Crying	293	5.2	2.6	21.7	< .01
Scatter	292	37.7	41.7	78.2	< .01
Rapport	290	16.1	13.5	26.1	*ns*
Hyperactivity	292	7.3	7.9	8.7	*ns*
Dependency on Mother	287	10.5	10.5	4.3	*ns*
Emotional Stability	287	16.2	8.4	30.4	< .01

ences are statistically significant, again at the .01 level, except for rapport, hyperactivity, and dependency.

Data on 600 more infants re-examined by another examiner are available as well as findings at 7 years of age. If this additional material were examined, the associations just demonstrated might very well be more firmly established and "softer" items of behavior for which no differences are demonstrated in the 300 cases might become evident. A significant relation with 40 week neurologic status is demonstrated in those areas where there are significant differences in behavior between the premature and the control infants. These differences between the prematures and controls are apparently being maintained even at 7 years of age (Harper, 1959). Since there was more neurologic abnormality in the premature than in the control infants at 40 weeks of age, we can expect that if the number of cases is large enough the association between 40 week neurologic status and later behavior can be demonstrated.

We will close this discussion by mentioning briefly some steps for future investigation suggested by the data and the theoretical discussion. We have made a beginning on the analysis of the items going into developmental assessment methods. Only the recently developed giant electronic digital computers can handle the massive amount of information available, and we are at present working on developing a program for a matrix of 250 by 250 items. This is a formidable task in itself, but when it is completed it would permit us to select the most predictive items and weight them objectively, rather than use the more difficult clinical judgments necessary at present. A very simple matter is the examination of the data already extant on the larger number of children both at 3 years and at 7 years of age. A final point is to approach the behavior of older children from the developmental viewpoint and construct measures which will distinguish between normal and abnormal neurologic integration. More can be done in this regard in infancy. Our tools beyond the period of infancy are apparently even less adequate.

SUMMARY

In the evaluation of deviations from normal neuropsychologic functioning in infancy, several points must be borne in mind. First, damage to the brain is usually diffuse and leads to a variety of disabilities. Second, there is a continuum of damage which ranges from minimal to severe, and there is a quantitative difference in the abnormal patterns over this range. The point at which the prognosis no longer is for full compensation, as evaluated by the standard neurologic examination in later life, is probably when "cerebral palsy" or other severe deviations become qualitatively different from the rest of the continuum. Third, there is an intimate interrelation between neurologic integrity and maturational level, and the diagnosis that is made on the basis of the abnormal behavior observed is dependent on the age and maturity of the infant. Development is an orderly process and the criteria for deviation from normal are well enough specified so that the behavior to be expected at a later stage can be predicted from that observed at earlier ages.

Data are presented which indicate that the behavioral approach, using the Gesell Developmental and Neurologic Examination, is a valid and reliable procedure capable of eliciting distinctive patterns in those areas of behavior in which abnormalities occur. If we wish to detect minimal degrees of abnormality in older children, we must use developmental concepts in forging additional diagnostic tools.

REFERENCES

GESELL, A., & AMATRUDA, CATHERINE S. *Developmental diagnosis*. New York: Paul B. Hoeber, 1954.

HARPER, P. A. Observations on growth and development of prematurely born children during the first seven years. In *Maryland Child Growth and Development Institute, June 1–5, 1959*. Baltimore: State Department of Health, 1959. Pp. 77–82.

KNOBLOCH, HILDA, & PASAMANICK, B. The syndrome of minimal cerebral damage in infancy. *J. Amer. med. Ass.*, 1959, *170*, 1384–1387.

KNOBLOCH, HILDA, & PASAMANICK, B. Environmental factors affecting human development, before and after birth. *Pediatrics*, 1960, *26*, 210–218.

MASLAND, R. L. Summary Session of the Ninth International Congress of Pediatrics: Neurology. *J. Pediat.*, 1960, *57*, 103–105.

13

Assessment of Infant Personality

Bettye M. Caldwell

In this article Caldwell analyzes the methodological and theoretical problems of research on infant personality. She stresses the active infant, the importance of studying the social climate with a particular focus on the mother-child diad, and direct observation as the most desirable research technique.

Research relating to infancy seems to be enjoying a vigorous renaissance period. This new cycle has something of a reputation to keep, for it is doubtful that more exemplary specimens of behavioral research can be found than some of the studies concerned with infancy published during the thirties and early forties. This generalization is particularly true with respect to studies of the neonate. One has only to read the early reviews by Hurlock (1933) and Richards and Irwin (1934) to be impressed with the quality and scope of the work done during the period.

With the exception of a few experiments relating to patterns of emotional development, little of the infancy research of the earlier period dealt with what would be considered personality variables. During the present era, however, personality development has been a topic of major concern. This has been expressed mainly in the form of investigations of different patterns of early infantile experience and their effects upon concurrent and future personality.

Convictions about the importance of the infancy period for personality development are found in persons of widely different theoretical orientations and with different standards about acceptable methods by which "facts" may be admitted to the corpus of scientific knowledge. Support for the assumption of an association has been sought in relation to the impact of highly specific events occurring within the infancy period (e.g., scheduled vs. demand feeding) and in relation to more global variables (e.g., warmth of the mother-child relationship, or adequacy of environmental stimulation).

This conviction about the importance of the infancy period for personality development has been difficult to extinguish, perhaps simply because it seems logical to people accustomed to thinking in terms of the continuity of life events. Its resistance to extinction may also be a function of the intermittent rather than the regular reinforcement it has received in terms of evidence from human subjects. In his well-known review of studies relating infantile experience to personality, Orlansky (1949) concluded with a note of pessimism resulting from the failure of most of the studies to show any clear-cut relationships. Since that time many additional studies dealing with the same problem, mostly of retrospective design, have appeared in the literature and have frequently reported negative or inconclusive findings. However, a vigorous and posi-

---

From *Merrill-Palmer Quarterly of Behavior and Development*, 1962, *8*: 71–81.

tive counterpoint for this little fugue has come from two sources: (1) a great wealth of clinical literature (Bowlby, 1952; Ribble, 1943; Spitz, 1946), stressing the critical nature of infantile experiences; and (2) an increasing and tantalizing accumulation of animal studies (Harlow, 1958; Thompson, 1955) which, almost without exception, have shown that certain modifications of the species-modal pattern of interindividual or individual-environment relationships can have long-lasting and possibly irreversible effects upon complex forms of behavior.

In the present discussion I shall try to formulate my suspicions that we are maneuvering ourselves into a *cul de sac* with respect to significant personality data relating to infancy and that the absence of appropriate methods has little to do with it. Rather it seems to this author that personality research in infancy has suffered from two related shortcomings: (1) failure to adapt techniques useful with older subjects and to incorporate methodological advances which have been made in recent years, and (2) premature closure with respect to variables which might profitably be investigated.

METHODOLOGICAL PROBLEMS

Problems of Sampling

Although much of our infancy research seems to have been carried out by methodologically sophisticated researchers, relatively little attention has been paid to sampling problems. We continue to work with what Kessen (1960) has called "unsampled" samples. In large measure this is related to the availability of subjects and has led to a concentration of attention upon the neonate. While research on the neonate is justified and important, it should hardly be regarded as adequate coverage for the entire infancy period. Indeed, there are in some respects fewer similarities between the neonate and the two-year-old than there are between the two-year-old and the adult. To offer as a defense the fact that older infants are difficult to obtain as subjects is not adequate.

Even within the heavily researched newborn period, however, there are sampling problems which have been largely ignored. In the first place, the available studies have dealt with hospital populations only. Although it is known that this in itself indicates a biased sample, recognition is seldom made of this fact. Furthermore, little concern has been shown for drawing cases at random from this particular sub-population. Rather, subjects used in a given project are more likely to be infants delivered by a particular obstetrician or cared for by a cooperative pediatrician or located in a nursery supervised by a sympathetic nurse. Implicit in this lack of regard for rigorous sampling procedures seems to be the notion that during the early weeks or months of life, infants are more similar than different and for many research purposes one is as good as another. Yet behavior research during infancy consists largely of a search for measurable responses which will show individual differences and thus serve as bases for prediction of later response modes. Such differences in a variety of response measures pertinent to personality research have been shown, and it seems unwise to ignore fundamental rules about sampling procedure in the selection of subjects.

The problem of obtaining representative samples of newborns does not seem quite as great as that of sampling behaviors occurring during the total range of what is generally considered the infancy period. Most of the more famous longitudinal studies (Jones & Burks, 1936; Stone & Onque, 1959) conducted over the past 30 years did not begin data collection until after the infancy period. Gesell and his co-workers (1934) have reported on the development of children throughout the infancy range, and, of course, the existing baby scales were generally standardized on infants at all ages up to two years. The work of Escalona and Leitch (1952), designated as non-normative, dealt with infants between four and 32 weeks of age, and there are several relatively young longitudinal studies in which data relating to the entire infancy period are being collected. But taken all together, these current studies probably cover only a few hundred babies.

If we accept the premise that the goal of behavioral research dealing with any age period is to enable us to predict subsequent behavior or development, then our willingness to accept if not prefer the neonatal period is surprising. For, as Richmond and Lipton (1959) have pointed out, "Studies of newborn infants can only be oriented physiologically since data collection must necessarily be physiologic (including motor behavior) or biochemical." The distance between the essentially physiological (though behavioral) variables identified in some of the careful studies of the newborn and subsequent variables generally classified as personality measures is undoubtedly great and involves many intervening steps, each of which can influence the subsequent behavior. Thus while predictions about later personality based on data collected during the neonatal period might be dazzling, they are in some ways a bit presumptuous at this stage of our knowledge. The human neonate has been compared to the opossum, which, at birth, just barely has what it takes to get from uterus to pouch. Whether a human newborn could make such a hazardous journey is perhaps questionable!

The informational lacunae relating to the middle and later months of the infancy period—during which many of the presumed critical events for almost any personality theory are likely to occur—suggest that much of our research on infant development should be centered in pediatric rather than educational settings. Perhaps through our increasingly efficient public health machinery some of these sampling problems can be remedied.

One final comment about the sampling problem seems in order. This is the observation that historically we have been content with frightfully small samples in much of the research dealing with infancy. While this tendency to work with small N's undoubtedly attests to the difficulties involved in finding infants for such research, it may also hint at the earlier-mentioned assumption that differences among infants are not as great and as significant as among older subjects. Although a prosaic one, inadequate sample size is surely one of the more remediable methodological problems associated with the study of infancy.

Techniques of Assessment

In personality research with infants, the past decade has been characterized by increasing neglect of situational methods of assessment. This may reflect the wariness of investigators to set up experimental situations which deviate even slightly from current ideas, however imperfectly validated, about ideal infant care. Perhaps as scientists we are all emotionally committed to the notion of the psychological vulnerability of the human infant, just as most new parents are to its physical frailty. Not since the forties—probably not since Spitz's report on hospitalism (1946) received wide distribution—have experimental studies of social isolation, such as those of the Dennises (1951) and Arsenian (1943), been found in the literature. Gardner and Pease (1958) use a 60-second separation of mother and child with preschoolers as a situational test, but the child is actually alone for only 15 seconds of that period.

Whether such experimental timidity is related more to fear of adverse consequences to the infant than the desire to escape opprobrium from colleagues can only be a matter for conjecture. Yet degrees of such isolation are probably repeated several times daily in millions of homes, as, for example, when a conscientious mother justifies letting her baby "cry it out" in his crib in order to keep from "spoiling" him, or perhaps has coffee in the apartment next door while her prematurely awakened infant tries unsuccessfully to announce that his nap is over. Both selection and avoidance of procedures are governed by the current state of knowledge in any field, and we are perhaps more alert to possible deleterious consequences when working with infants than with any other group of subjects. Yet it is worth considering whether situational procedures which investigate relevant issues about infant experiences and their consequences would be as psychologically damaging as the experience of reporting fear or feelings about his parents might be to a four-year-old, or suffering an experimentally-induced failure might be to a ten-year-old. . . .

This wariness about manipulating variables has led to an unhealthy reliance on second-hand methods of securing information. The current reliance upon the interview as the main source of data in personality research dealing with infancy has reached disturbing proportions. You have a problem to investigate, so you schedule-ize an interview and investigate it. Thomas *et al.* (1960) have an interesting rationale for their choice of the interview in preference to standard laboratory or testing procedures. They assert that interview-elicited data will reflect characteristic behavior, whereas possibly atypical and unrepresentative behavior will emerge from the artificial laboratory situation. Such a point of view seems to disregard the distortions possible in a parental report, even when every effort is made to limit the reporting to behavioral observables and to minimize interpretation. It would seem that the investigator needs to be especially alert to possible confounding of results when the same interview is used to measure both the independent and the dependent variable, e.g., parent practices and personality of the child. Current reporting does not automatically solve the problem of second-hand information, and I know of no evidence to demonstrate less distortion in prospectively, as compared to retrospectively, obtained interviews.

If we are going to use them essentially as psychometric instruments, then comparable standards of psychometric reliability and validity should be demanded of research interviews. Response sets can and should be systematically investigated. Operation of a social desirability set is implied in the statement by the Maccobys: "Validity studies all suggest that when people are interviewed directly concerning behavior about which there is a strong expectation of social approval or disapproval, and in which there is considerable ego-involvement, respondents tend to err in the direction of idealizing their behavior" (1954, p. 482). And a recent study by Joan and William McCord (1961) showed many differences between theoretically significant family life variables in two large, carefully equated groups, one of which was studied by observation and the other by interview. The picture of family life obtained from the interview group was closer to the cultural stereotype of the happy family than was that based on direct observation, e.g., parents did not reject the children, children did not reject the parents, interparental esteem was high, discipline less punitive, etc. . . .

It would seem that personality research relating to the infancy period can escape the traps of second-hand reporting and anamnestically acquired data by making greater use of the kinds of parameters which can be directly observed and measured experimentally and relating these to subsequent personality measures. A prime example of this can be found in learning studies. As demonstrated in countless papers in the literature, here is a phenomenon which can be investigated reliably in early infancy. Furthermore, techniques are available for studying different manifestations of the phenomenon throughout the life span. Yet learning studies have not traditionally been concerned with changes over time or with the developmental implications of typical learning patterns in individuals.

For example, there are enough studies in the literature to demonstrate that conditioning can occur in infancy, although there are differences of opinion as to the basal age. But so far as the present author knows, we have no data on the subsequent developmental careers of children who do and who do not condition during early infancy. Is the child who can be conditioned at an early age more vulnerable to environmental vicissitudes? And what about the infants who cannot be conditioned? Will they subsequently be more resistant to chance pairings of stimuli likely to result in the development of a conditioned response? As early as 1929 Gesell (1929), in an attempt to combat what he considered an over-zealous acceptance of the presumed importance of conditioning for structuring behavioral morphology, suggested that possibly maturational factors helped to make an infant secure against either favorable or unfavorable conditioning. The imprinted fowl has, in some respects, acquired a maladaptive response. So also might an infant who, on the basis of a few chance pairings, learns to make sucking movements every time the phone rings.

. . . We need to know more about how early conditionability relates to subsequent measures of instrumental learning in the child. We are beginning to collect data on some of these problems in our laboratory. A companion approach would call for more intensive study of the kinds of learning situations, either accidental or deliberate, set up for the infant in the home by the parents.

. . . Learning studies in infancy could easily be structured as investigations of the development of an extremely important parameter for personality study. Such research in the past has concentrated on increased understanding of the process of learning rather than on an increased understanding of the infants studied. It would seem that here is an area in which much research is needed and where many of the methodological problems have already been dealt with, if not solved.

SELECTION OF VARIABLES

Premature Closure Related to Theory

This will not be a popular point, as it is currently more fashionable to stress the importance of theory-derived research (McCandless & Rosenblum, 1952). With respect to personality theory relating to infancy, psychoanalytic theory has unquestionably been the dominant influence. The possibility has to be considered that this preoccupation with problems relevant to one or another aspect of psychoanalytic theory has trapped our attention and prevented our observing other behaviors which might have even greater relevance for developmental theory. Consider, for instance, how preoccupied we have been with feeding and weaning and toilet training—*ignes fatui* of socialization theory, one might call these events. And, for a variety of reasons, the theory remains essentially untouched by the accumulation of related empirical data. This is not the optimal relationship between theory and research evidence; either they nourish one another or their mutual obligation is not fulfilled.

To say that there has been little feedback to the theory from the empirical studies is not to imply that the theory has not undergone significant modification. But it does not seem that much of this restructuring has resulted from the empirical studies *per se*. Small wonder, when one reflects on the contradictory and inconclusive nature of the empirical data. Rather than being affected by results from empirical studies, it seems that psychoanalytic theory has been modernized by certain basic changes in techniques for securing information. The early hypotheses came from adult anamnestic material. But many contemporary child analysts are engaged in projects which involve observations, not reconstruction, of behavior of normal and disturbed children. And it is significant to note that different variables find their way into analytic reports when live, not anamnestic, childhood is under study.

Modification of psychoanalytic theory (or methods of data collection) is only incidental to the point here—namely, that there is a need for many more studies of the human infant in his natural habitat, the equivalent of the field studies of biology. The methods so carefully worked out and applied by Barker and Wright (1955) and associates need little or no modification in order to be appropriate for the study of the infant. Because of ethical considerations, as mentioned previously, our ability to manipulate variables which are possibly relevant for personality development during infancy is limited. This means that we need to make every effort to capitalize on experiments in nature which permit us to avoid manipulation and control of variables by capitalizing on selection. Of all the currently used methods of personality research dealing with older children and adults, those involving direct observation of behavior would seem to require the least modification in order to be suitable for use with infants. If we are concerned with the "natural history of child behavior" (Wright, 1960), ecological studies of infants should not be overlooked.

A One-Sided Approach, or the Problem of the Dichotomized Organism

In this modern era of behavior study, terms like "biosocial" and "psychosomatic" sound forth with the dull thud of a stale cliché. An investigator who would deny such an orientation would probably be hard to find. But in terms of the methods employed and the parameters chosen for investigation, how much research exemplifies this orientation? It would seem that, starting from either side of this false dichotomy, one can only reach a point of no return, always some distance away from the opposite goal. Personality theorists whose orientation is essentially social generally pay tribute to such concepts as organismic or temperamental variables but then design research that ignores them. Likewise, the biological personality theories of Gesell (1934), Shirley (1933), and others—in spite of their lip-serviced reminders that organism and environment cannot be completely separated—leave one with the feeling that there is nothing to do but watch and wait. . . .

Genuine adherence to, not mere verbal acceptance of, a bio-social approach would reflect a significant advance in personality research. An interesting historical example of a non-practiced preachment dealing with this point can be found in a 1929 article by Weiss (1929). After stating his advocacy of a bio-social approach, he then offers a comment on methodology which rather effectively excludes it. In describing their intensive 24-hour observations of the behavior of newborns throughout their hospital stay, he reports, "The infant is under continuous observation except for those periods which are spent with the mother and during which he is bathed and fed" (1929, p. 457). That is, the infant was observed at all moments except those likely to be crucial for establishing patterns of social interaction. Theoretical positions and methodological choices do not always harmonize. But it seems a correct generalization that here is an area in which the renaissance "Infant" and the old "Infant" should be different. The relevant social diads of mother-infant, father-infant, etc., must not be investigated in a biological vacuum but in terms of those aspects of bodily functioning which can be reliably measured during infancy and which will mediate the performance of the infant in diadic behavioral sequences.

A bio-social research commitment takes the baby out of the aseptic atmosphere of the stabilimeter into the warm though perhaps contaminating arms of "mother." Personality research that adequately covers the mother-infant diad is likely to depend largely upon observation of live behavior rather than report or anamnesis and to consider the infant as a stimulating as well as a passively absorbing organism. If some of these problems of method and variable selection will be confronted, it should be safe to predict that the new "Infant" will grow and thrive and not, like his sibling of thirty years ago, develop theoretical marasmus and waste away.

REFERENCES

ARSENIAN, JEAN M. Young children in an insecure situation. *J. abnorm. soc. Psychol.,* 1943, *38,* 225-249.

BARKER, R. G., & WRIGHT, H. F. *Midwest and its children: The psychological ecology of an American town.* New York: Row, Peterson & Co., 1955.

BELL, R. Q. Relations between manifestations in the human neonate. *Child Develpm.,* 1960, *31,* 463-477.

BOWLBY, J. *Maternal care and mental health.* Geneva: World Health Organization, 1952.

DENNIS, W., & DENNIS, SENA G. Development under controlled environmental conditions. In W. Dennis (Ed.), *Readings in child psychology.* New York: Prentice-Hall, 1951. Pp. 104-131.

ESCALONA, SIBYLLE, LEITCH, MARY, *et al.* Early phases of personality development: A non-normative study of infant behavior. *Monogr. Soc. Res. Child Develpm.,* 1952, *17,* No. 1, vi+72.

GARDNER, D. B., & PEASE, D. The use of situational tests with preschool children. *Nursery Educ.,* 1958, *14,* 18-20.

GESELL, A. Maturation and infant behavior patterns. *Psychol. Rev.,* 1929, *36,* 307-319.

GESELL, A., & THOMPSON, HELEN. *Infant behavior: its genesis and growth.* New York: McGraw-Hill, 1934.

HARLOW, H. F. The nature of love. *Amer. Psychologist,* 1958, *13,* 673-685.

HURLOCK, ELIZABETH B. Experimental studies of the newborn. *Child Develpm.,* 1933, *4,* 148-163.

JONES, MARY C., & BURKS, BARBARA S. Personality development in childhood. *Monogr. soc. res. Child Develpm.,* 1936, *1,* No. 4, vi+205.

KESSEN, W. Research design in the study of developmental problems. In P. H. Mussen (Ed.), *Handbook of research methods in child development.* New York: Wiley, 1960. Pp. 36-70.

MACCOBY, ELEANOR E., & MACCOBY, N. The interview: a tool of social science. In G. Lindzey (Ed.), *Handbook of social psychology.* Cambridge, Mass.: Addison-Wesley, 1954. Pp. 449-487.

McCANDLESS, B. R., & ROSENBLUM, S. Psychological theory as a determiner of experimental pattern in child study. *Rev. educ. Res.,* 1952, *22,* 496-525.

McCORD, JOAN, & McCORD, W. Cultural stereotypes and the validity of interviews for research in child development. *Child Develpm.,* 1961, *32,* 171-185.

ORLANSKY, H. Infant care and personality. *Psychol. Bull.,* 1949, *46,* 1-48.

PEASE, D., & GARDNER, D. B. Research on the effects of noncontinuous mothering, *Child Develpm.,* 1958, *29,* 141-148.

RIBBLE, MARGARET A. *The rights of infants: early psychological needs and their satisfaction.* New York: Columbia University Press, 1943.

RICHARDS, T. W., & IRWIN, O. C. Experimental methods used in studies on infant reactions since 1900. *Psychol. Bull.,* 1934, *31,* 23-46.

RICHMOND, J. B., & LIPTON, E. L. Some aspects of the neurophysiology of the newborn and their implications for child development. In Lucie Jessner & Eleanor Pavenstedt (Eds.), *Dynamic psychopathology in childhood.* New York: Grune & Stratton, 1959. Pp. 78-105.

SHIRLEY, MARY. *The first two years: A study of twenty-five babies.* Vol. III. *Personality manifestations.* Minneapolis: University of Minnesota Press, 1933.

SPITZ, R. A. Hospitalism. In O. Fenichel, *et al.* (Eds.), *The psychoanalytic study of the child.* Vol. I. New York: International Universities Press, 1946. Pp. 113-117.

STONE, A. A., & ONQUÉ, GLORIA C. *Longitudinal studies of child personality.* Cambridge, Mass.: Harvard University Press, 1959.

THOMAS, A., CHESS, STELLA, BIRCH, H., & HERTZIG, MARGARET E. A longitudinal study of primary reaction patterns in children. *Comprehensive Psychiatry,* 1960, *1,* 103-112.

THOMPSON, W. R. Early environment: its importance for later behavior. In P. H. Hoch & J. Zubin (Eds.), *Psychopathology of childhood.* New York: Grune & Stratton, 1955. Pp. 120-139.

WEISS, A. P. The measurement of infant behavior. *Psychol. Rev.,* 1929, *36,* 453-471.

WRIGHT, H. F. Observational child study. In P. H. Mussen (Ed.), *Handbook of research methods in child development.* New York: Wiley, 1960. Pp. 71-139.

14

The Measurement of Maternal Care

Harriet L. Rheingold

How does the researcher measure the mother-child relationship? One approach is to investigate the mother's attitudes and behavior and then to determine their effects on the infant. Rheingold attacks the problem by utilizing direct observation of the caretaking situations in home and institution and analyzing the effects upon a limited number of infant activities.

Maternal care is commonly assumed to have an effect upon infant behavior, but experimental verification of the assumption is still only sketchy. Of general notions about possible relationships between maternal care and infant behavior there is no dearth, nor are suitable measures of infant behavior lacking. Rather, it is measures of maternal care which are needed before statements concerning the effects of maternal care upon infant behavior can be drawn in a more precise, and hence testable, form.

Questionnaires, interviews, and case histories have customarily been used to obtain assessments of the mother's personality traits and attitudes and to obtain reports of her child rearing practices. But personality assessments and retrospective accounts are several steps removed from the actual care the infant receives (Bell, 1958) and, furthermore, do not easily lend themselves to the kinds of exact measures needed for statements of association.

The purpose of this paper is to describe a method of measuring maternal care in terms of the actual operations the mother performs in caring for the infant. It is a method which gives direct access to many of the important environmental events of the infant's life and thus establishes more closely the rationale for relating maternal care to infant behavior; at the same time it should be possible to gain clues to the mother's attitudes and personality traits from her caretaking behavior. The method employs the technique of time-sampling maternal care as it occurs in its natural setting. As demanded by the technique (Arrington, 1943), both the items of behavior to be recorded and the time schedule of observations are exactly specified in advance. The details of the method will be illustrated by an account of a study in which the care given three-month-old infants in their own homes was compared with care given by an institution. The results obtained will be presented and then used to evaluate the method's success in providing the needed measures.

METHOD

Subjects

Ten infants were the subjects of study; five of them were home infants, and five were residents of an institution. The mean age of the home *S*s was 3.4 months (range, 3.2 to 3.7), of the institution *S*s, 3.5 months (range, 3.2 to 3.8). Age was the only criterion used for choosing the *S*s in the institution. In addition to age, home *S*s had to be "only" infants of parents of better than average socioeconomic status, requirements which it was expected would increase the probability of finding a difference between home and institutional care. The home *S*s were located by word of mouth, and requests for cooperation were made by telephone. It was necessary to ask only six mothers to find the five *S*s; one refused on the plea of personal illness. All *S*s were females except for one male in the institutional group.

Description of Homes and Institution

The *S*s and their parents lived without other relatives in small houses or apartments which were simply but imaginatively furnished. Each *S* had his own room.

The *S*s in the institution, resident since birth, lived in a large room which housed 19 infants. The caretakers included the Sister in charge, who was a trained nurse, attendants, and volun-

---

From *Child Development*, 1960, *31:* 565–575. Reprinted with permission of the Society for Research in Child Development.

teers. In general, the caretakers were sensitive to the needs of the infants, while the infants themselves appeared to be well-developed, active, and responsive.

Procedure

The Checklist. On the basis of observation and trial (Rheingold, 1956) a checklist composed of 42 items was drawn up, 30 for mothering activities and 12 for infant activities (Table 1 gives the items). Of the 30 mothering items six recorded the location of *S* in his environment, four the number of people in his room, and four the number of people within six feet of him.

The remaining 16 items recorded activities more closely related to mothering. "Caretaking" was an all-inclusive item which was checked if at the moment of observation someone was doing something for *S* while in proximity to him. "*N* of caretakers" recorded the number of different caretakers who cared for *S* during the eight-hour period of observation. The last 14 items were sufficient to cover all the caretaking operations performed by people in caring for three-month-old infants at the level of analysis selected for study here. Although all items were defined in detail, the behaviors included in them need no comment, with a few exceptions. "Talks" was checked when the caretaker talked to anyone and not specifically to *S*; her speech could be thought of as providing auditory stimulation for the infant. "Plays" included both social play of caretaker and infant *and* caretaker's initiating infant's play with toys. The item "Other" had to be used very infrequently; it covered such caretaking activities as giving *S* vitamins and weighing him.

It should be noted that the caretaking activities were not considered as mutually exclusive. Rather, all the activities occurring at any moment of observation were checked; thus, a caretaker could be bathing the infant, looking at his face, and talking to him on the same observation.

Although the purpose of the study was to measure the caretakers' activities, a number of infant activities which served to delineate maternal care were also included.

Time of Observation. For each *S* a period of eight hours was sampled according to the following schedule: *O* looked at *S* for a full second and then recorded on the checklist what he was doing, if someone was caring for him, the nature of the caretaking act, where he was, and how many other people were in his environment. *O* could take as long as 14 seconds to complete the record, but in any case she did not look up again until the 15th second when she again observed *S* for a full second. In this fashion, four observations were made in each minute for a total of 10 minutes. After a break of five minutes another sample of 10 minutes was observed. Thus, the first 10 minutes of each quarter hour were sampled for four consecutive hours, that is, from 9 A.M. to 1 P.M. on one day and from 1 P.M. to 5 P.M. on the next day, to make a total of eight hours. There were three *O*s in all, but only a pair for each infant, one observing the first four hours, and the other, the second four.

Observers' Behavior. The *O*s endeavored to put the mothers and the attendants at their ease by being friendly and by conspicuously making the infant the focus of observation. The *O*, for example, stayed in the room with *S*, whether he was asleep or awake, and continued to make observations. The readiness with which *O*s were accepted was reassuring, as was the extent to which people went about the business of caring for *S as though* they were not being observed. Mothers, on questioning subsequent to the period of observation, said that *O*'s presence inhibited only their more "foolish" behavior with the infant, e.g., using nonsense syllables, singing, etc. The effect of *O*'s presence will be raised again in the Discussion.

Scoring. The score was the frequency of occurrence of each item over the eight-hour period for each subject. From these scores a mean score for each item was derived for each group of five *S*s.

Since 40 observations were made during each 10-minute period, the total number of observations for the eight-hour period was 1280. Dividing the mean scores by 1280 could give percentages of occurrence; and, since the observations were systematically distributed throughout time, the mean score could also be translated into hours and minutes.

Observer-Agreement. To check the reliability of the observations, *O*s made independent simultaneous observations during 64 10-minute periods. Thirtv-five of the periods occurred before the study began, the rest at intermediate points as the study progressed, in order to insure continuing agreement. The observations were made on seven *S*s, four in the institution and three in homes, none of whom were subjects in the study.

The amount of observer-agreement for each item was calculated by two different methods. In the first, a percentage of agreement was obtained by dividing the number of agreements by the combined number of agreements and disagreements. Here an agreement was a tally identical in both records; that is, both *O*s agreed not only in *identifying* the item but also in the *timing* of the item; agreement on the absence of an activity was *not* counted as an agreement. The median percentage of agreement for all items was 89.8 (range 16.7 to 100.0). As one might expect, there was some tendency for *O*s to agree more often on frequently occurring items. (The rank-order correlation between percentage of agreement and frequency of items was .71, $p < .01$.)

In the second method, observer-agreement was calculated by correlating the frequencies obtained for each item. The median of the product-moment correlations was .97 (range .35 to 1.00).

On the basis of these findings the method was judged to have sufficient observer-agreement to warrant a measure of confidence in the results obtained by its use.

RESULTS

Table 1 gives the means and the ranges of the scores for the items of maternal care and infant activities.

. . . Two-tailed tests seemed indicated because what was at test was the power of a method to differentiate between two groups.

Caretaking

Amount of Caretaking. The range of frequencies for "Caretaking" in the two groups not only did not overlap but were in fact widely separated. Home infants were cared for on 37 per cent of the observations to the institution infants' 8 per cent. Thus, they received care 4.5 times as often.

Number of Caretakers. Most commonly the mother was the only caretaker for the home infant during the period of observation. In one home, however, a father, and in another, an aunt, gave some caretaking. The institution infant, on the other hand, was cared for by six different caretakers during the eight-hour period of observation.

Caretaking Activities. Table 1 shows that the home infant received more of every caretaking activity than the institution infant, with the exception of "Dresses." While the home infant received care 4.5 times as often, the differences in amounts of different kinds of caretaking were even greater. He was looked at 4.9 times as often, fed on 5.4 times as many observations, held 6.2 times as often, and talked to 9.5 times as often. Some less frequent kinds of caretaking occurred even more often in the home. For example, the home infant was patted 13.1 times as often, played with 14.8 times as often, rocked 16.9 times as often, and shown affection 18.0 times as often.

The results also reveal that in most caretaking activities the differences were so great that there was no overlapping of scores between the two groups. All the differences were statistically significant, except those for "Shows affection," "Bathes," "Dresses," "Rocks," and "Other."

Amount vs. Kind. The difference between care in the home and in the institution appeared to be not in the kinds of caretaking activities (as defined here) but rather in the amounts of the different kinds. First, every activity performed by a mother in the home was performed, however infrequently, in the institution. Second, there were marked similarities between home and institution in the caretaking activities occurring most frequently. For the home, the activities in order of magnitude were "Holds," "Talks," "Talks to," "Feeds," and "Looks at face"; in the institution the order was "Holds," "Feeds," "Looks at face," "Talks to." And lastly, the ratio of each caretaking activity to the total amount of caretaking was similar for home and institution. Differences, therefore, tended to be differences in amounts and not in kinds.

Location of Infant. The infant in the institution spent more time in his crib than the home infant, but the difference was not significant. When the home infant was out of his crib, he was most often in his mother's arms; in contrast, the institution infant, when out of his crib, was most often seated in a swing. "Other location" occurred significantly more often for the home infant and included play pen, parents' bed, and sofa. The home infant also spent much more time out of his own room. The comparison, however, is not entirely appropriate: the institution infant's room was often as large as, or larger than, the entire apartment of the home infant; and different parts of his room were used for different purposes and hence were furnished differently.

Number of People in the Environment. The institution infant spent less time alone, with no one in his room, than the home infant. Furthermore, there were more often two or more adults in the

Table 1: *Means and ranges of material care and infant activity N of Infants in Each Group = 5 N of Observations for Each Infant = 1280*

Item	Home		Institution	
	Mean	Range	Mean	Range
**Caretaking	475.2	347– 617	105.8	74– 144
** *N* of caretakers	1.4	1– 2	6.0	5– 8
** Looks at face†	140.2	109– 173	28.4	6– 36
** Talks	55.8	26– 105	7.8	0– 19
** Talks to infant	166.2	51– 310	17.0	12– 22
** Pats	26.2	9– 44	2.0	1– 4
Shows affection	3.6	0– 14	0.2‡	0– 1
* Plays	11.8	4– 25	0.8‡	0– 4
** Holds	285.4	196– 394	45.6	11– 70
** Diapers	37.4	27– 57	5.6	1– 10
Bathes	35.2	0– 103	8.8	6– 11
** Feeds	154.4	72– 198	28.6	4– 43
Dresses	11.2	0– 28	16.2	6– 31
** Adjusts position	6.8	0– 11	2.2	1– 4
Rocks	37.2	0– 138	2.2	0– 4
Other	14.2	8– 20	11.4	3– 18
Location of Infant				
In crib	787.0	667– 908	908.4	699–1081
** In arms	258.6	179– 391	46.2	8– 76
** Seated	0.0	——	260.0	128– 530
Bathinette	86.0	27– 139	41.2	19– 70
Other location	148.2	2– 327	24.2	0– 112
** Out of own room	304.4	153– 391	43.0	21– 77
N Adults in Room				
** 0	690.0	466– 899	263.0	178– 352
1	534.8	372– 814	464.4	416– 568
** 2	34.8	0– 136	246.0	182– 329
** 3+	20.0	0– 99	304.2	210– 361
N Adults within 6 Ft.				
** 0	755.5	652– 919	1062.6	1022–1093
** 1	492.2	361– 621	161.0	130– 213
* 2	12.8	0– 38	45.0	33– 68
3+	19.5	0– 78	11.4	5– 18
Infant Activities				
Awake	847.4	565–1081	928.6	647–1042
Vocalizes	54.2	22– 81	38.4	1– 76
Vocalizes to mother	9.6	1– 30	3.0	0– 5
Protests	14.4	5– 21	20.6	1– 59
Fusses	31.4	14– 46	43.6	19– 71
Cries	23.6	2– 64	14.8	0– 47
Finger in mouth	103.8	0– 312	69.0	10– 111
** Bottle in mouth	54.4	0– 146	298.8	170– 435
* Plays with toy	62.8	9– 168	18.2‡	0– 91
Plays with hand	8.4	2– 20	41.6	1– 117
* Plays with dress	0.0	——	8.4	2– 20
Bursts of activity	32.6	13– 78	46.0	18– 92

† *Caretaking items not mutually exclusive.*
‡ *Mean based on one case.*
* *Difference between groups significant at .05 level, two-tailed test.*
** *Difference between groups significant at .01 level, two-tailed test.*

room with him. But, because the room in the institution was so large, these people were often at a distance from him. Thus, on the one hand, the institution infant saw and heard more people upon occasion, while, on the other, he more often had no adult in his *near* environment. The home infant had far fewer persons in his environment, but these few (most often, the mother, of course) were more often with him.

Variability of Caretaking Activities. Not only did the means for maternal care differ in the two groups, but, as has been noted earlier, the variances also differed widely. On 14 of the 16 mothering items, the ratio of home variances to institution variances was greater than 1.0, the median ratio being 19.5. Only for the items, "*N* of caretakers" and "Other," were the institution variances greater.

The differences in the variances present more than a statistical problem. They suggest that the care home infants receive may vary a great deal from one home to another, while the care given in an institution may be more similar from infant to infant. If this conclusion is warranted, and if maternal care does affect infant behavior, then one might expect home infants to differ more among themselves, institutionalized infants to differ less. Of course, repeated measures of maternal care in the same homes would be needed before definitive statements could be made about the variability of care between individual mothers. One might also find by means of repeated measures in the institution whether some infants elicit more care than others.

Infant Activities

Contrary to the findings on caretaking activities, measures of infant activities showed much smaller differences between the two groups, as may be seen in Table 1. In only three activities, "Bottle in mouth," "Plays with toy," and "Plays with dress," were the differences large enough to achieve statistical significance. All three were activities which reflect differences in the caretaking practices of the two environments rather than in infant behavior; the institution infant's bottle was propped; he had few, if any, toys in his crib; and he wore a short, loose shirt. If "Plays with toy" is combined with "Plays with hand" and "Plays with dress," the difference between the groups becomes negligible. One can conclude that, in the absence of toys, the infants in the institution played an equal amount, but with their hands and their clothes, that is, with what the environment offered. . . .

Variability of Infant Activities. It will be recalled that the variability in caretaking activities was much greater in the home than in the institution. One might therefore have expected a similar difference in variability to appear in the infant activities. The data, however, did not support the expectation, for the variances of infant activities in the two groups of infants were similar.

Part vs. Whole Measures

Now that differences in maternal care in institution and home have been demonstrated for samples of eight-hour periods, it may be asked whether shorter samples would have yielded differences of the same magnitude. To answer the question the results for the first four hours (9 A.M. to 1 P.M.) were compared with those for the second four (1 P.M. to 5 P.M.). . . . A four-hour sample of maternal care then would have been adequate to obtain most of the differences yielded by the eight-hour sample.

CONCLUSION AND DISCUSSION

Two separate topics may be considered here, the results of the comparison of maternal care in two environments and an evaluation of the method. The first will be discussed only briefly because the comparison was made primarily to illustrate the method.

Results of the Study

The results showed that the home infant had many more contacts with a mother than the institution infant with a caretaker. Not only was more done for him but he was brought into contact with more things. He had a richer life, a more complex environment. At the same time, one can speculate, his life was probably more the same from day to day. To the extent that a mother gave care in her own way, his life would lack the kind of complexity brought about by many different and frequently changing caretakers.

While the contrasts in care were noteworthy, so too were the similarities. The results showed, for example, that caretaking activities were similar in kind and in order of frequency, whether in home or in institution.

That no differences of statistical significance were found in the infant activities (that is, in those not dependent on caretaking practices) raises several questions. It may be that, infrequent as the institutional care was, it was nevertheless sufficient at this age to arouse and maintain behavior similar to that shown by home infants. Indeed, the very infrequence and variableness of the institution's schedule of reinforcement, in contrast to the home's more regular schedule, may have accounted for its potency in shaping and maintaining selected acts of behavior. Interesting as such speculations are, however, it must be pointed out that the method was designed to measure maternal, rather than infant, behavior.

To look, further, for association between caretaking activities and infant behaviors in the results obtained is tempting but, once again, caution is in order; had differences been found, they could not have been attributed to differences in maternal care alone.

Evaluation of the Method

The results of the study demonstrate that a set of maternal care variables can be specified, defined, and measured. Clear and unequivocal differences in these variables were found in two environments; as a consequence, exact, quantitative statements can now be made about the extent of the differences.

Limitations of the Method. Implicit in the method are characteristics which some might regard as limitations. The presence of the observer is one of these characteristics, and, admittedly, it may alter the mother's usual behavior. Although the observer can by his deportment do much to reduce the extent and the nature of the distortion, this limitation, more generally, will need to be considered afresh for each task set the method. In the present case, for example, there is no reason to believe that all the mothers were similarly stimulated by the observer's presence to give *more* care than they usually did or all the caretakers in the institution to give *less* care. It need only be mentioned in this connection that mechanical recorders may pose still other problems.

It should also be obvious that the measures yielded by the method are discrete and discontinuous; the 14 seconds between the observations are silent. Thus, the method offers no information about the sequence of stimulus and response between mother and infant.

Research Implications. With exact measures of maternal care at hand it should be possible to reproduce, experimentally, environments offering certain kinds and amounts of maternal care, thereby gaining a measure of control.

Without revision the method could be used, also, to answer a number of questions. How does maternal care vary from mother to mother? How does it differ with the socioeconomic status of the home? What are the characteristics of maternal care in different cultures? How does maternal care vary with the age or birth order of the child?

Given only the requirement that the items of behavior and the schedule of observations be defined in advance of measurement, the method will be seen to possess considerable flexibility. In the present study only one of the possible sets of maternal care variables has been used; other sets of variables, for example, one employing finer discriminations of behavior, come to mind.

Finally, it is probably not enough to demonstrate that maternal care can be measured. In the end, the measures will gain validity only to the extent to which they can be shown to enter into predictable relationships with infant behavior.

SUMMARY

The purpose of this paper was to describe a method of measuring maternal care in terms of the actual operations a mother performs in caring for an infant. The method employed the technique of time-sampling, with both the items of behavior to be recorded and the time schedule of observations specified in advance. The method was illustrated by comparing care given three-month-old infants in their own homes with care given in an institution.

On a checklist containing 30 items of mothering activities and 12 items of infant behavior, the observer recorded what both mother (or caretaker) and infant were doing on any second of observation. Observations were made every 15th second for the first 10 minutes of every consecutive quarter hour for a period of eight hours. A record for an infant therefore consisted of 1280 observations systematically distributed over eight hours.

The results showed that the home infants received care 4.5 times as often as infants in an institution; the differences in some specific caretaking operations were even greater. The differences, however, tended to be differences in amounts of caretaking and not in kinds (as defined here). Measures of infant activities, on the other hand, yielded only small and, in general, not reliable differences.

The results demonstrate that a set of maternal care variables can be defined and measured. As a consequence, it may now be possible to formulate and test quantitative statements about the relationships between maternal care and infant behavior.

REFERENCES

ARRINGTON, RUTH E. Time-sampling in social behavior. *Psychol. Bull.,* 1943, *40,* 81–124.

BELL, R. Q. Retrospective attitude studies of parent-child relations. *Child Develpm.*, 1958, *29,* 323–338.

RHEINGOLD, HARRIET L. The modification of social responsiveness in institutional babies. *Monogr. Soc. Res. Child Develpm.*, 1956, *21,* No. 2 (Serial No. 63).

RHEINGOLD, HARRIET L. A method for measuring maternal care. *Amer. Psychologist,* 1958, *13,* 319. (Abstract)

15

Development of a Parental Attitude Research Instrument

Earl S. Schaefer and Richard Q. Bell

Schaefer and Bell, like Rheingold, conducted their study as staff members of the National Institute of Mental Health. Their aim was to develop an attitude instrument that would indicate how mothers felt about family life and children. Theoretically, these attitudes would then be revealed in the behavior of the mother toward the child and would thus influence the child's development. This study reports only the first step, the development of such an instrument; the article by Shaw and Dutton (49) illustrates one of its applications.

This paper will present (a) the thesis that parental attitudes toward child-rearing and the family life are an important influence in the personality development of the child, (b) a review of objective studies of parental attitudes toward child rearing, (c) a rationale for the development of homogeneous measures of concepts concerning parental attitudes and a description of the process of concept development, (d) the results of an analysis of item and scale efficiency, and (e) correlations of the attitude measures with education.

The current extensive interest in the influence of the parent-child relationship upon the personality development of the child arose primarily through psychotherapeutic work with adults and the clinical studies of families by child guidance clinics. Many of the earlier studies have been reviewed by Symonds (1939) and Radke (1946). Orlansky's review (1949) of literature relevant to the hypothesis that types of infant care determine personality structure suggested the need for an emphasis upon the total pattern of experiences of children rather than upon specific infant care practices in the study of environmental influences upon personality development. Symonds (1939) also concluded that the quality of the parent-child relationship is of crucial importance. Mead's theory of the genesis of the self is a concise statement of this point of view: "The individual experiences himself as such, not directly, but only indirectly, from the particular standpoints of other individual members of the same social group, or from the generalized standpoint of the social group as a whole to which he belongs" (1934, p. 138). Since the most extensive and intensive social interactions of the child during crucial developmental stages occur within the family and especially with the mother, the mother-child relationship would be of major importance in personality development.

Recent trends in personality theory provide a rationale for the conceptualization and quantification of maternal attitudes. Bronfenbrenner (1953) has noted a convergence upon a set of related concepts which imply a tendency to orient and act with respect to objects. In action theory these tendencies have been called need-dispositions while similar concepts have variously been called dispositions, dispositional tendencies, needs, belief-value matrices, and dynamic systems. These concepts include emotional, cognitive, and motivational elements and imply stable genotypical personality characteristics. Let us assume that such need-dispositions are related to the type of relationship which a woman develops with her child. Development of measures of those components of her personality which are relevant to her role as a mother would permit prediction of her behavior with her child and the future personality adjustment of the child. The hypothesis that such need-dispositions can to some degree be determined by measuring the attitudes of parents toward child-rearing and the family is the motivation for this study.

The importance of maternal attitudes to the development of the child has been generally accepted. Bettelheim (1952) has pointed out that if parents who rigidly attempt to impose accelerated developmental goals upon children are told that early toilet-training is undesirable they may delay such training but may continue to

---

From *Child Development,* 1958, *29:* 339–361. Reprinted with permission of the Society for Research in Child Development.

accelerate the child in talking, reading, and other important areas of development. This example illustrates that an underlying attitude may influence a great variety of parent behaviors. Many researchers have implicitly assumed that objective attitude measures could give valid information on such parental attitudes. The potential usefulness of such objective measures justifies their development. When such measures are available, their construct validity can be objectively determined.

If objective attitude measures can be demonstrated to be indices of certain patterns of maternal behavior, attitude measurement would be a valuable supplement to interview and observation as methods for the study of maternal behavior. The Maccobys' discussion of the relative merits of interviews and questionnaires is relevant to the question of validity; they state: "It has been widely assumed that the interview is superior to the questionnaire in many ways, and should be used where resources permit. Certainly the interview *must* be used at exploratory stages; for the later stages of research, the meager evidence so far available does not support the superiority of the interview. . . . It is entirely possible that the interview is superior for some subject matters, the questionnaire for others" (1954, p. 483). If methods for coding interviews were well developed, ample time for thorough interviews were available, and well trained interviewers could be obtained, it seems likely that interviews would give more valid information. However, the speed, efficiency and objectivity of the questionnaire make it a desirable method at this stage of research on parent-child relationships.

REVIEW OF PREVIOUS STUDIES

Turning to the empirical evidence for the validity of this approach to the study of parent-child relationships, we find a number of researches which have studied relationships between parental attitudes and the personality adjustment of children. Stogdill (1936) was among the first to attempt the objective measurement of parental attitudes with two questionnaires which measured attitudes toward parental control and attitudes toward social behavior of children. Read (1945) studied the relationship of these variables with a behavior inventory for nursery school children by comparing the responses of 32 parents who were tested with these measures with the rated behavior of their 20 children. Although she concluded that liberalism in views on parental control is related to child behavior, correlations between scores on parental attitude scales and summarized ratings were not significant when computed from tabular data presented in the research report. Anderson (1946) reported a study completed by Miles (1945) in which it was found that parents' attitudes are related to the leadership status and social acceptance of the child by the school group. He reported that subscales measuring overprotection, dominance by the parent, and encouragement of social development are related to the child's social adjustment. The effect of a significant difference in education between parent groups was not partialled out. Shoben (1949) administered an inventory of attitudes toward child-rearing to 50 mothers of children in mental hygiene clinics and juvenile courts and 50 mothers of normal children. Significant correlations between three parental attitude scales and children's adjustment were shown on cross-validation with comparable groups. Scales measuring maternal dominance, possessiveness, and ignoring were approximately equal in power of differentiation. Partial correlations computed from data available in the published report indicate that a significant difference in educational level between parents of problem and nonproblem children did not materially affect the validity coefficients obtained.

Radke (1946) constructed scales of parental attitudes for the areas of autocratic or democratic control of the child, amount and area of restrictions on the child, amount of freedom, severity or mildness of punishment, rapport between parent and child, the relative responsibility of the parents for the child's discipline, and sibling compatibility or incompatibility. No reliabilities for these scales were given but the intercorrelations of the measures were low. Radke stated that relationships between these home atmosphere scales and child behavior ratings by preschool teachers suggested that unfavorable conduct of children is related to autocratic, restrictive, and severe discipline. These conclusions were based on finding 11 differences significant at or beyond the .05 level, where seven would be expected by chance, in 140 comparisons of 20 different behavior ratings of preschool children and seven attitude scales administered to their parents. The extreme homogeneity of the group of parents on education and other characteristics may have operated to depress relationships.

. . . Harris, Gough, and Martin (1950) also tested the relationship between a priori scales which measured authoritarianism, permissiveness, parent-child integration, parental rigidity, and good judgment (the responses judged by the writers to be most desirable) and an 18-item attitude scale toward Negroes which was given to the children of these parents. The number of items differentiating high and low prejudice groups differed only slightly from what would be expected by chance at the significance levels cited. Empirical parental attitude scales constructed from the data were not cross-validated.

Shapiro (1952) conducted a study of correlates of parental attitudes on child-rearing with other attitudes and objective measures on parents. He found that attitude items which approve the free expression of a child's desires are related to objective attitude measures of "Radicalism" with parents who are less restrictive being more radical in their political opinions. Items which he designated as "Analytical" dealing with anal, oral, and sibling-rivalry situations were less related to these attitudes. He concluded that attitudes toward child-rearing are related to other social attitudes and that different types of parental attitude items are differentially related to other personality measures of the parent. Block (1955), in a more comprehensive study of restrictive attitudes toward child-rearing, found a scale of paternal restrictiveness significantly correlated with 122 of 498 other tests and measures which were included in a personality assessment of 100 military officers. Paternal restrictiveness was also related to Q-sort personality evaluations which were made by eight staff psychologists who had observed the subjects during the procedures of the assessment program. A comparison of the 20 most restrictive and the 20 most permissive subjects revealed that "Restrictive fathers tend to be constricted, submissive, suggestible individuals with little self-assurance. Permissive fathers were evaluated as self-reliant, ascendant, and functioning effectively. . . ." Restrictive attitudes were also positively related to measures of ethnocentricism and fascism. . . .

Croty (1957) failed to find any correlations in excess of .25 between measurements of achievement motivation, obtained from a psychological insight questionnaire administered to 95 male high school students, and attitude measures obtained from their mothers. The attitude questionnaire, Final Form IV of the Parental Attitude Research Instrument (PARI), which is reported in this paper, was mailed to the mothers and a return of 78 per cent was obtained.

Klebanoff (1957) administered Form IV of PARI to 15 mothers of schizophrenic children, 15 mothers of retarded or brain injured children, and 26 mothers of normal children. Significant differences in scales related to an overpossessiveness factor appeared between mothers of normals and mothers of schizophrenics as well as between mothers of normals and mothers of retarded or brain injured children. Differences remained after statistical control of education differences between groups. No differences could be demonstrated between mothers of schizophrenic children and mothers of mentally retarded or brain damaged children on factor scores, but differences between groups on individual scales were not tested.

Mann (1957) also administered Final Form IV of PARI as well as questionnaires developed by Shoben (1949) and Mark (1953) to mothers of children with cerebral palsy and closely matched mothers of normal children. Mothers of normals scored significantly higher on the Ignoring scale from Shoben, but differences on the Possessive and Dominant scales were not significant. Mothers of children with cerebral palsy scored significantly higher (showed more agreement) on the PARI scales Seclusiveness, Strictness, Intrusiveness, Acceleration, Encouraging Verbalization, Equalitarianism, and Comradeship and Sharing. Mothers of normals scored significantly higher on Marital Conflict, Irritability, Rejection of the Homemaking Role, and Avoidance of Communication. Eleven of the 23 PARI scales yielded differences significant at or beyond the .05 level.

This review of studies of parental attitudes toward child-rearing reveals great interest in the hypothesis that objectively measured attitudes toward child-rearing are significantly related to personality development of children. Most of the studies quoted here are interpreted by their authors as supporting the hypothesis. Despite this consensus the research findings do not seem to be conclusive due to problems of matching experimental and control groups, the use of ad hoc scales without adequate test development, the attribution of significance to a small number of items out of a larger group under circumstances where such items might well have reflected chance relationships, the marginal significance of the findings of several studies, as well as the general absence of cross-validation of findings of marginal significance.

Several of the studies reviewed, as well as a study by Wiley (1950), have developed measures of attitude concepts such as parental dominance, control, or restrictiveness; possessiveness; ignoring; and parent-child rapport or integration.

The conceptual schemes have often been too limited, concept definition has often been too vague, or adequate test development has been neglected. The review of the literature suggested a need for the development of a comprehensive conceptualization and quantification of this domain which would provide a research instrument for future studies.

CONCEPTUAL ANALYSIS OF PARENTAL ATTITUDES

Both clinical studies, in which various syndromes of parental attitudes have been related to differential personality development of children, and the statistical theory of differentiation and prediction suggest the need for development of measures of specific attitude concepts composed of homogeneous groups of items. Theoretically, such an inventory would be more useful than a test composed of individual items or one which would give only one score of pathogenicity of parental attitude. . . . Guttman's (1950) criticisms of heterogeneous composite scores also suggest that the use of only one composite score in the area of parental attitudes would tend to reduce differentiating power of a parental attitude research instrument and would obscure interpretation of results.

With these considerations in mind a conceptual analysis of the domain of parental attitudes was carried out as a basis for developing measures of attitudes. The relatively heavy saturation of statistically significant items in the studies by Mark (1953) and Shoben (1949) prompted sorting of their item pools by three clinical psychologists into subscales or groups of items which seemed psychologically homogeneous. In conference these conceptual analyses were discussed and a group decision was made to use the concepts given in Table 1. This conceptual scheme was used to classify all items of the two studies into one of the categories given. This analysis, although based upon varying numbers of items in the various categories, strongly suggested that some categories have higher discriminating power than others.

It also became evident, in the process of sorting, that items which make what seem to be either purely factual statements or which state attitudes that would be generally endorsed by psychologists as expressing approved attitudes toward child-rearing do not differentiate effectively between parents of normals and parents of maladjusted children. This is in keeping with the report of Stogdill (1936) who noted that often parents will agree with apparently contradictory attitudes if one statement expresses what appears to be a healthy attitude while the other states its converse. Harris, Gough, and Martin (1950) also report that items stating healthy atti-

Table 1: *Conceptual analysis of items from Mark and Shoben studies*

Parental Attitudes	No. of Items differentiating beyond .05 level of significance		No. of Items not differentiating		Per Cent of items differ-entiating
	Mark's Study	*Shoben's Study*	*Mark's Study*	*Shoben's Study*	*Combined*
Strictness	3	6	0	0	100
Intrusiveness	3	2	0	0	100
Martyrdom	2	1	0	0	100
Breaking the Will	2	5	1	0	88
Infantilization	6	5	2	0	87
Deification	6	10	4	0	80
Suppression of Aggression	1	2	1	0	75
Child-dominated Home	1	9	0	4	71
Nonpunishment	3	2	2	0	71
Suppression of Sexuality	6	4	1	3	71
Ignoring the Child	4	5	4	2	60
Avoidance of Tenderness	3	0	1	1	60
Absolute Authority of Parent	2	13	7	5	56
Social Isolation of Child	2	6	5	3	50
Acceleration	4	1	6	0	45
Statement of Fact	0	0	8	6	00
Expression of Positive Attitude	0	0	4	14	00
Unclassified Items	19	14	26	25	39

tudes are typically poor discriminators. We found that scales which state approved attitudes toward child-rearing typically have very poor reliabilities because there is a strong tendency for all persons to agree. As a result there is little variability and poor discrimination on so-called healthy scales.

This finding led us, in constructing a set of scales for this domain, to concentrate our efforts upon the development of the more differentiating scales which state attitudes contrary to the usually approved child-rearing opinions. It was soon found that tests composed entirely of such scales left many mothers very dissatisfied. They frequently expressed the feeling that they did not see any items with which they could agree and therefore could not indicate their own ideas. To correct this we included in our own inventory a number of scales which we should expect, from previous studies, to have relatively low reliabilities and poor discrimination. These so-called "Rapport" scales are Equalitarianism, Expression of Affection, Comradeship and Sharing, Encouraging Verbalization, and Autonomy of the Child.

Our procedure in developing additional scales was as follows. After developing a concept on the basis of theory or formulating a concept from the writings of others, we attempted an operational definition by writing items designed to measure the concept. Scales which were thus developed were tried out in short exploratory forms on samples of 25 mothers. Items were eliminated if they did not reveal variation in attitudes or did not vary with other items which were used to define the scale. Promising items were noted, and new items were written based upon hypotheses about the successful items. Usually two or three successive trial forms, each of which utilized the insights derived from previous trial forms, were used in developing an item pool which was then standardized on samples of 100. This procedure permitted a process of hypothesis development and testing which maximized the amount of new information which could be used in developing an effective item pool, and did not impose statistical rigor at the level of hypothesis formation. Use of successive samples, and confirmation of small sample work with larger samples in the final forms, supported the hypothesis that this process resulted in improvement in operational definition of the concepts.

The results of this sequential experimentation compared with results which would have been obtained by a simple elimination of items from an original item pool is shown in the development of the scale Harsh Punishment. The reliability or internal consistency of the original pool of 10 items for the scale Harsh Punishment was .12, which could have been improved to the level of .15 by elimination of four items. However, by writing new items to follow up promising leads this scale was finally developed into an eight-item form with a reliability of .82. Improvements in scale reliability were frequently rather dramatic following detection of the psychological basis of the success of discriminating items.

The selection of concepts for development was based upon a search of the literature for hypotheses as to which areas would give additional information relevant to parent-child relationships. As a result of the hypothesis that information on marital adjustment would assist greatly in understanding the home environment of the child, several scales were developed which measured relevant concepts. The scale Marital Conflict contains items which were intended to reveal the existence of tensions indirectly through obtaining the respondents' judgments about how much quarreling and dissension might be expected in marriages in general. The scale Rejection of the Homemaking Role contains items which state the unhappiness of a woman at being shut up in a home, and her dissatisfaction with the duties of caring for the home and children. The scale Irritability was developed around items which indicate that children "get on a woman's nerves" and that any woman would "blow her top" frequently in the difficult job of managing a home. This scale may be related to several sources of tension but it seems feasible that it would also reflect tensions in the marital relationship. The scale Inconsiderateness of the Husband contains items which state how inconsiderate husbands in general are. It is intended to permit projection of feeling about the spouse. Ascendancy of the Mother, which measures the extent to which the woman plans to dominate the family, was also developed to investigate a possible source of marital conflict.

The two scales, Encouraging Verbalization and Avoidance of Communication, were written to determine whether the parent would permit or encourage the child to talk about his anxieties, conflicts, hostilities, and disagreements with parental policies. Ignoring the Baby was designed to determine how much affection and attention the mother thought was appropriate for the care of the infant. These scales which measure attitudes toward communication and toward care of the infant hopefully should measure the warmth and permissiveness of the parent-child

Table 2: *Education of PARI samples contrasted with 1950 census data**

	Primiparae $N=200$ *Per Cent*	Multiparae $N=200$ *Per Cent*	1950 Census *Per Cent*
Grammar School	2	2	13
Some High School	10	13	25
High School Graduate	50	49	42
Some College	23	24	12
College Graduate	13	10	6
No information	2	2	2

* *1950 census data on education of all white females, urban and rural, within the age range of 18 to 40 years.*

relationship. The scale Fear of Harming the Baby was developed out of interest in clinical reports of the eruption of unconscious hostile feelings toward their infants in disturbed mothers. Seclusion of the Mother was designed to measure the social adjustment of the mother, since it would presumably influence the socialization of the child. Dependency of the Mother is an attempt to determine the extent to which the mother looks to others for security and support. Fostering Dependency was designed to measure another aspect of overprotection or overpossessiveness. Harsh Punishment attempts to determine the degree to which the mother feels punishment is an effective method of influencing and controlling children. Excluding Outside Influences, which is designed to measure family ethnocentrism, may be related to parental control and authoritarian attitudes.

Abdication of the Parental Role reflects an attitude that parents can do relatively little with children and that most parental efforts are wasted. This scale, with Rejection of the Homemaking Role and Irritability, was included in an effort to determine the extent to which the mother rejects or ignores her children. We felt it important to explore this area since the items utilized in previous studies seemed to be more relevant to clinical descriptions of parents of withdrawn or schizophrenic children than to parents of aggressive or delinquent children. Demand for Activity items measure the extent to which the parent feels children should be kept busy and rushing at all times.

Although no systematic effort was made to develop generalizations concerning successful item-writing, it was frequently found that vague items which presumably enhance projective responses could be used to explore difficult areas. Rationalizations used by parents of disturbed children were used with the hope that psychopathogenic parents would agree while the critical judgment of more normal parents would cause them to reject the statement. An example is the item, "Children need some of the natural meanness taken out of them," in the scale Breaking the Will. The apparent truism seems to give the respondent with a latent but defended attitude a feeling that it may be socially acceptable to agree. Cliches, colloquialisms, and conventional affect-laden phrases were often useful. Examples of these are phrases in the scale Irritability such as "will get on any woman's nerves," "can't stand their children a moment longer," "it's natural for a woman to blow her top." An example of attaching the aura of self-righteousness to an item is, "A wise parent will teach a child early just who is boss." Such techniques were used in item writing with the hypothesis that thresholds for item acceptance would be changed differentially in normal and deviant parents. No empirical data are available, however, to test this hypothesis.

THE SAMPLES

During two different two-month periods women were tested within one to four days following delivery of their babies at a military hospital in the metropolitan area of Washington, D.C. Mothers who were not sufficiently recovered from sedation at the time of test administration, mothers with less than an eighth grade education, and non-English speaking mothers were excluded. Age and educational levels of these groups are reported in Tables 2 and 3. These data are reported as a partial assessment of the representativeness of these groups. Test instructions* were given on the ward, each mother

---

* See Appendix B—*Editor's note.*

Table 3: *Ages of PARI samples contrasted with 1952 vital statistics**

	PARI Samples		1952 Vital Statistics		
Age *Years*	Primiparae *N*=200 *Per Cent*	Multiparae *N*=200 *Per Cent*	Age *Years*	Primiparae *Per Cent*	Multiparae *Per Cent*
–18	9	2	–19	25	4
19–21	36	12	20–24	44	26
22–24	34	25			
25–27	9	15	25–29	20	33
28–30	5	15	30–34	8	23
31–33	2	13	35–	3	14
34–36	3	10			
37–	0	7			
No information	2	1			

* *Class intervals cannot be compared directly since PARI data were obtained in class intervals which differed from vital statistics class intervals.*

Table 4: *PARI scale correlations with educational level*

Attitude Scale	*Primiparae*	*Multiparae*
	(*N* = 100)	(*N* = 100)
Equalitarianism	–.11	–.02
Nonpunishment	–.03*	–.32
Avoidance of Tenderness	–.09	–.20
Suppression of Aggression	–.33	–.34
Breaking the Will	–.20	–.42
Autonomy	–.31	–.12
Ignoring the Child	–.06	–.15
Strictness	–.13	–.24
Intrusiveness	–.30	–.40
Expressing Love and Affection	–.02	–.06
Infantilization	–.29	–.42
Suppression of Sexuality	–.33	–.48
Acceleration of Development	–.31	–.46
Comradeship and Sharing	–.05	.14
Deification of the Parent	–.26	–.52
Martyrdom	–.38	–.42
Encouraging Verbalization	.09	.05
	(*N* = 96)	(*N* = 97)
Seclusiveness of the Mother	–.30	–.22
Harsh Punishment	–.29	–.27
Dependency of the Mother	–.28	–.11
Fear of Harming the Baby	–.27	–.27
Fostering Dependency	–.33	–.31
Marital Conflict	–.35	–.17
Irritability	–.22	–.26
Excluding Outside Influences	–.25	–.15
Rejection of the Homemaking Role	–.07	–.18
Avoidance of Communication	–.30	–.20
Ignoring the Baby	–.11	–.03
Ascendance of the Mother	–.21	–.19
Inconsiderateness of the Husband	–.20	–.20
Approval of Activity	–.08	–.24
Abdication of the Parental Role	–.33	–.20

* *Positive correlation should be interpreted as meaning a person with a high education level agrees with the attitude expressed by the scale.*

working individually while remaining in her bed. Trial Form III was given to 100 freshman and sophomore student nurses. Final Form IV,* which was used to determine the reliability of the five item scales, was given to a group similar to the samples for Trial Forms I and II. . . .

Although predictive validity studies have not yet been done on the PARI, the studies surveyed in the review of literature could be cited as evidence supporting the concurrent validity of this general approach to the study of parent-child relationships. The content validity or the adequacy of the concepts of parental attitudes toward child-rearing can be determined by attempting to add concepts which would add new variance to measurement of this domain. Adequate evidence of the construct validity of these scales must be based upon research on various correlates of these measures.

CORRELATIONS OF SCALES WITH EDUCATIONAL LEVEL

Previous investigations (Davis & Havighurst, 1946; Maccoby *et al.*, 1954) have found that differences in attitudes toward child-rearing are related to the social class level of the respondents. A recent investigation of the interrelationships of various indices of social class level led to the conclusion that level of educational achievement is highly related to other indicators of socioeconomic level (Kahl & Davis, 1955). These attitude scores were therefore correlated with self-reports of number of years of school completed successfully. The findings, which are reported in Table 4, seem to be congruent with the results reported by Maccoby *et al.* (1954) in that mothers of the lower educational levels tend to have less approved attitudes toward children. The correlations of education with attitude for the scales which state disapproved attitudes toward child-rearing range from −.03 to −.38 for primiparae with a mean of −.24, and from −.03 to −.52 for multiparae with a mean of −.27. Although many of these correlations are statistically significant at beyond the .05 level, the absolute magnitude of the correlations does not justify the interpretation of the variations in attitude as merely reflecting differences in educational level.

SUMMARY

Theories of parental influence upon development of children and a review of previous research on the relationship of parental attitudes to the personality adjustment of children suggested the need for the development of a set of homogeneous measures of parental attitudes. A set of 32 concepts were selected which were derived from previous studies and from a search of the literature on parent-child relationships.

Attitude scales of five to 10 items which gave satisfactory reliability for research purposes were developed with an iterative technique of attitude measurement. Many of these measures were found to be related to education, with mothers of higher education having more usually approved attitudes toward child-rearing. It is suggested that logically, psychologically, and empirically homogeneous scales of attitude toward child-rearing will be useful in investigating theories of the influence of maternal attitudes upon development of the child.

REFERENCES

ANDERSON, J. E. Parents' attitudes on child behavior: A report of three studies. *Child Develpm.*, 1946, *17*, 91–97.

BETTELHEIM, B. Mental health and current mores. *Amer. J. Orthopsychiatry*, 1952, *22*, 76–78.

BLOCK, J. Personality characteristics associated with fathers' attitudes toward child rearing. *Child Develpm.*, 1955, *26*, 41–48.

BRONFENBRENNER, U. Personality. In C. P. Stone & D. W. Taylor (Eds.), *Annual review of psychology*. Stanford, Calif.: Annual Reviews Inc., Vol. 4, 1953. Pp. 157–182.

CROTY, C. J. Maternal attitudes and their relation to son's achievement motivation. Unpublished master's thesis, Catholic University, 1957.

DAVIS, A., & HAVIGHURST, R. J. Social class and color differences in child-rearing. *Amer. sociol. Rev.*, 1946, *11*, 698–710.

GUTTMAN, L. Relation of scalogram analysis to other techniques. In S. A. Stouffer *et al.* (Eds.), *Measurement and prediction*. Princeton, N.J.: Princeton University Press, 1950. Pp. 172–212.

HARRIS, D. B., GOUGH, H. B., & MARTIN, W. E. Children's ethnic attitudes: II. Relationship to parental beliefs concerning child rearing. *Child Develpm.*, 1950, *21*, 169–181.

KAHL, J. A., & DAVIS, J. A. A comparison of indexes of socioeconomic status. *Amer. sociol. Rev.*, 1955, *20*, 317–325.

KLEBANOFF, L. B. Attitudes of mothers of schizo-

---

* See Appendix A—*Editor's note*.

phrenic, brain injured and retarded, and normal children. Unpublished doctoral dissertation, Boston University, 1957.

MACCOBY, ELEANOR E., GIBBS, PATRICIA K., & the staff of the Laboratory of Human Development, Harvard University. Methods of child-rearing in two social classes. In W. E. Martin & Celia B. Stendler (Eds.), *Readings in child development*. New York: Harcourt, Brace, 1954. Pp. 380–396.

MACCOBY, ELEANOR E., & MACCOBY, N. The interview: A tool of social science. In G. Lindzey (Ed.), *Handbook of social psychology*. Vol. I. Cambridge, Mass.: Addison-Wesley, 1954. Pp. 449–487.

MANN, V. D. A study of the attitudes of mothers of cerebral palsied children toward child adjustment. Unpublished doctoral dissertation, American University, 1957.

MARK, J. C. The attitudes of the mothers of male schizophrenics toward child behavior. *J. abnorm. soc. Psychol.*, 1953, *48*, 185–189.

MEAD, G. H. *Mind, self, and society*. Chicago: University of Chicago Press, 1934.

MILES, KATHERINE A. Relationships between certain factors in the home background and the quality of leadership shown by children. Unpublished doctoral dissertation, University of Minnesota, 1945.

ORLANSKY, H. Infant care and personality. *Psychol. Bull.*, 1949, *46*, 1–48.

RADKE, MARIAN J. Relation of parental authority to children's behavior and attitudes. *University of Minnesota Inst. Child Welf. Monogr.*, 1946, No. 22.

READ, KATHERINE H. Parents' expressed attitudes and children's behavior. *J. consult. Psychol.*, 1945, *9*, 95–100.

SHAPIRO, M. B. Some correlates of opinions on the upbringing of children. *Brit. J. Psychol.*, 1952, *43*, 141–149.

SHOBEN, E. J., JR. The assessment of parental attitudes in relation to child adjustment. *Genet. Psychol. Monogr.*, 1949, *39*, 101–148.

STOGDILL, R. M. The measurement of attitudes toward parental control and the social adjustment of children. *J. appl. Psychol.*, 1936, *20*, 359–367.

SYMONDS, P. M. *The psychology of parent-child relationships*. New York: Appleton-Century, 1939.

WILEY, J. H. A scale to measure parental attitudes toward certain aspects of children's behavior. Unpublished doctoral dissertation, University of California, 1950.

APPENDIX A

SAMPLE COPY OF FINAL FORM IV (PARI)[1]

Inventory of Attitudes on Family Life and Children

Read each of the statements below and then rate them as follows:

A	a	d	D
strongly agree	mildly agree	mildly disagree	strongly disagree

Indicate your opinion by drawing a circle around the "A" if you strongly agree, around the "a" if you mildly agree, around the "d" if you mildly disagree, and around the "D" if you strongly disagree.

There are no right or wrong answers, so answer according to your own opinion. It is very important to the study that all questions be answered. Many of the statements will seem alike but all are necessary to show slight differences of opinion.

---

[1] Permission is granted to anyone to reproduce this material with proper acknowledgment, without permission of the authors or of the Society for Research in Child Development.

	Agree		Disagree	
1. Children should be allowed to disagree with their parents if they feel their own ideas are better.	A	a	d	D
2. A good mother should shelter her child from life's little difficulties.	A	a	d	D
3. The home is the only thing that matters to a good mother.	A	a	d	D
4. Some children are just so bad they must be taught to fear adults for their own good.	A	a	d	D
5. Children should realize how much parents have to give up for them.	A	a	d	D
6. You must always keep tight hold of baby during his bath for in a careless moment he might slip.	A	a	d	D
7. People who think they can get along in marriage without arguments just don't know the facts.	A	a	d	D
8. A child will be grateful later on for strict training.	A	a	d	D
9. Children will get on any woman's nerves if she has to be with them all day.	A	a	d	D
10. It's best for the child if he never gets started wondering whether his mother's views are right.	A	a	d	D
11. More parents should teach their children to have unquestioning loyalty to them.	A	a	d	D
12. A child should be taught to avoid fighting no matter what happens.	A	a	d	D
13. One of the worst things about taking care of a home is a woman feels that she can't get out.	A	a	d	D
14. Parents should adjust to the children some rather than always expecting the children to adjust to the parents.	A	a	d	D
15. There are so many things a child has to learn in life there is no excuse for him sitting around with time on his hands.	A	a	d	D
16. If you let children talk about their troubles they end up complaining even more.	A	a	d	D
17. Mothers would do their job better with the children if fathers were more kind.	A	a	d	D
18. A young child should be protected from hearing about sex.	A	a	d	D
19. If a mother doesn't go ahead and make rules for the home the children and husband will get into troubles they don't need to.	A	a	d	D
20. A mother should make it her business to know everything her children are thinking.	A	a	d	D
21. Children would be happier and better behaved if parents would show an interest in their affairs.	A	a	d	D
22. Most children are toilet trained by 15 months of age.	A	a	d	D
23. There is nothing worse for a young mother than being alone while going through her first experience with a baby.	A	a	d	D
24. Children should be encouraged to tell their parents about it whenever they feel family rules are unreasonable.	A	a	d	D
25. A mother should do her best to avoid any disappointment for her child.	A	a	d	D
26. The women who want lots of parties seldom make good mothers.	A	a	d	D
27. It is frequently necessary to drive the mischief out of a child before he will behave.	A	a	d	D
28. A mother must expect to give up her own happiness for that of her child.	A	a	d	D
29. All young mothers are afraid of their awkwardness in handling and holding the baby.	A	a	d	D
30. Sometimes it's necessary for a wife to tell off her husband in order to get her rights.	A	a	d	D
31. Strict discipline develops a fine strong character.	A	a	d	D
32. Mothers very often feel that they can't stand their children a moment longer.	A	a	d	D
33. A parent should never be made to look wrong in a child's eyes.	A	a	d	D
34. The child should be taught to revere his parents above all other grown-ups.	A	a	d	D
35. A child should be taught to always come to his parents or teachers rather than fight when he is in trouble.	A	a	d	D

36. Having to be with the children all the time gives a woman the feeling her wings have been clipped. A a d D
37. Parents must earn the respect of their children by the way they act. A a d D
38. Children who don't try hard for success will feel they have missed out on things later on. A a d D
39. Parents who start a child talking about his worries don't realize that sometimes it's better to just leave well enough alone. A a d D
40. Husbands could do their part if they were less selfish. A a d D
41. It is very important that young boys and girls not be allowed to see each other completely undressed. A a d D
42. Children and husbands do better when the mother is strong enough to settle most of the problems. A a d D
43. A child should never keep a secret from his parents. A a d D
44. Laughing at children's jokes and telling children jokes makes things go more smoothly. A a d D
45. The sooner a child learns to walk the better he's trained. A a d D
46. It isn't fair that a woman has to bear just about all the burden of raising children by herself. A a d D
47. A child has a right to his own point of view and ought to be allowed to express it. A a d D
48. A child should be protected from jobs which might be too tiring or hard for him. A a d D
49. A woman has to choose between having a well run home and hobnobbing around with neighbors and friends. A a d D
50. A wise parent will teach a child early just who is boss. A a d D
51. Few women get the gratitude they deserve for all they have done for their children. A a d D
52. Mothers never stop blaming themselves if their babies are injured in accidents. A a d D
53. No matter how well a married couple love one another, there are always differences which cause irritation and lead to arguments. A a d D
54. Children who are held to firm rules grow up to be the best adults. A a d D
55. It's a rare mother who can be sweet and even tempered with her children all day. A a d D
56. Children should never learn things outside the home which make them doubt their parents' ideas. A a d D
57. A child soon learns that there is no greater wisdom than that of his parents. A a d D
58. There is no good excuse for a child hitting another child. A a d D
59. Most young mothers are bothered more by the feeling of being shut up in the home than by anything else. A a d D
60. Children are too often asked to do all the compromising and adjustment and that is not fair. A a d D
61. Parents should teach their children that the way to get ahead is to keep busy and not waste time. A a d D
62. Children pester you with all their little upsets if you aren't careful from the first. A a d D
63. When a mother doesn't do a good job with children it's probably because the father doesn't do his part around the home. A a d D
64. Children who take part in sex play become sex criminals when they grow up. A a d D
65. A mother has to do the planning because she is the one who knows what's going on in the home. A a d D
66. An alert parent should try to learn all her child's thoughts. A a d D
67. Parents who are interested in hearing about their children's parties, dates and fun help them grow up right. A a d D
68. The earlier a child is weaned from its emotional ties to its parents the better it will handle its own problems. A a d D

69. A wise woman will do anything to avoid being by herself before and after a new baby. A a d D
70. A child's ideas should be seriously considered in making family decisions. A a d D
71. Parents should know better than to allow their children to be exposed to difficult situations. A a d D
72. Too many women forget that a mother's place is in the home. A a d D
73. Children need some of the natural meanness taken out of them. A a d D
74. Children should be more considerate of their mothers since their mothers suffer so much for them. A a d D
75. Most mothers are fearful that they may hurt their babies in handling them. A a d D
76. There are some things which just can't be settled by a mild discussion. A a d D
77. Most children should have more discipline than they get. A a d D
78. Raising children is a nerve-wracking job. A a d D
79. The child should not question the thinking of his parents. A a d D
80. Parents deserve the highest esteem and regard of their children. A a d D
81. Children should not be encouraged to box or wrestle because it often leads to trouble or injury. A a d D
82. One of the bad things about raising children is that you aren't free enough of the time to do just as you like. A a d D
83. As much as is reasonable a parent should try to treat a child as an equal. A a d D
84. A child who is "on the go" all the time will most likely be happy. A a d D
85. If a child has upset feelings it is best to leave him alone and not make it look serious. A a d D
86. If mothers could get their wishes they would most often ask that their husband be more understanding. A a d D
87. Sex is one of the greatest problems to be contended with in children. A a d D
88. The whole family does fine if the mother puts her shoulders to the wheel and takes charge of things. A a d D
89. A mother has a right to know everything going on in her child's life because her child is part of her. A a d D
90. If parents would have fun with their children, the children would be more apt to take their advice. A a d D
91. A mother should make an effort to get her child toilet trained at the earliest possible time. A a d D
92. Most women need more time than they are given to rest up in the home after going through childbirth. A a d D
93. When a child is in trouble he ought to know he won't be punished for talking about it with his parents. A a d D
94. Children should be kept away from all hard jobs which might be discouraging. A a d D
95. A good mother will find enough social life within the family. A a d D
96. It is sometimes necessary for the parents to break the child's will. A a d D
97. Mothers sacrifice almost all their own fun for their children. A a d D
98. A mother's greatest fear is that in a forgetful moment she might let something bad happen to the baby. A a d D
99. It's natural to have quarrels when two people who both have minds of their own get married. A a d D
100. Children are actually happier under strict training. A a d D
101. It's natural for a mother to "blow her top" when children are selfish and demanding. A a d D
102. There is nothing worse than letting a child hear criticisms of his mother. A a d D
103. Loyalty to parents comes before anything else. A a d D
104. Most parents prefer a quiet child to a "scrappy" one. A a d D
105. A young mother feels "held down" because there are lots of things she wants to do while she is young. A a d D
106. There is no reason parents should have their own way all the time, any more than that children should have their own way all the time. A a d D

107. The sooner a child learns that a wasted minute is lost forever the better off he will be. A a d D
108. The trouble with giving attention to children's problems is they usually just make up a lot of stories to keep you interested. A a d D
109. Few men realize that a mother needs some fun in life too. A a d D
110. There is usually something wrong with a child who asks a lot of questions about sex. A a d D
111. A married woman knows that she will have to take the lead in family matters. A a d D
112. It is a mother's duty to make sure she knows her child's innermost thoughts. A a d D
113. When you do things together, children feel close to you and can talk easier. A a d D
114. A child should be weaned away from the bottle or breast as soon as possible. A a d D
115. Taking care of a small baby is something that no woman should be expected to do all by herself. A a d D

APPENDIX B

INSTRUCTIONS TO SUBJECTS FOR PARI

"I am .... In cooperation with (hospital or school authorities) I am studying what (mothers) (women) think about how children should be brought up. A lot is written on this subject in various newspaper and magazine articles. Frequently these articles are not in agreement. I thought it would be a good idea to find out what (mothers) (women) themselves think. You can help in the study by passing on your own ideas. This is voluntary. If you would contribute your ideas I will leave the form with you and return for it in approximately 20 minutes. Be frank and give your own personal views regardless of what others may think. You do not need to give your name. I would, however, like you to fill out the questions on the last page for research purposes (age, education, etc.). So as not to use too much of your time we have a list of ideas which other mothers have contributed. You merely circle one of the four letters by each statement. Circle the large "A" if you strongly agree, the small "a" if you mildly agree, the small "d" if you mildly disagree, the large "D" if you strongly disagree. If you have any ideas which you feel should be included jot them down at the end. We would appreciate having them. Others who have given us their ideas say that it is best to work rapidly. Give your first reaction. If you read and reread the statements it tends to be confusing and you can't finish in the amount of time we have."

16

The Influence of the Let-Down Reflex in Breast Feeding on the Mother-Child Relationship

Niles Newton

This short paper illustrates the relationships between psychological and physiological variables within the mother and the effects of these variables upon the nursing situation for the infant. Breast feeding is thus seen as an interpersonal situation rather than a mechanical act. In examining this topic, Newton refers to various experimental studies with both infrahuman and human mothers.

How can anything as elemental as a physiological reflex influence something as complex as the mother-child relationship? There are two kinds of reflexes. Some of these, like the knee jerk and the salivary reflex, have little direct bearing on complex human interactions. But there is an important group of reflexes that are directly involved in interpersonal relations. These are the reflexes that help to regulate sexual and breast feeding behavior.

There are at least four reflexes involved in successful breast feeding. In the baby, the sucking reflex and the reflex that causes the head to turn in the direction of pressure help consummate the feeding act. In the mother, the nipple erection reflex and the let-down reflex are fundamental. We know little about the nipple erection reflex except that it exists and makes the nipple easier for the baby to grasp. However, we have recently learned quite a bit about the let-down reflex and its influence on mother and baby.

Farmers have long known something psychologists and physiologists have just discovered. It takes the cooperation of the cow to get milk out of the udder. That is why we have heard so much about contented cows. The udder can be full of milk, but if the cow is frightened or upset she will give little of her milk no matter how hard she is milked by hand or by milking machine.

The let-down reflex is involved. This is the way the let-down reflex acts (Newton & Newton, 1948). The nipple is sucked, and the sensory impulses involved in sucking are carried to the posterior pituitary gland which discharges oxytocin into the blood. The blood carries the oxytocin back down to the breast. The oxytocin causes the contractile tissue around the alveoli to contract. The alveoli are small grape-like structures that actually hold most of the milk. As they are squeezed together they push out the milk into the bigger ducts where it is available to the baby's sucking.

Although this mechanism has been worked out by detailed controlled experiments (Newton & Newton, 1948, 1950; Ely & Peterson, 1941; Peterson & Ludwick, 1942; Peterson, 1942), the let-down reflex is actually quite easy to observe in new mothers. In the first place, one breast often starts dripping milk while the baby is sucking on the other breast. Or if the let-down reflex is very strong and the breast is very full, sometimes more milk is pushed out than the baby can cope with. The baby pulls back its head and a stream of milk several inches long can be seen shooting from the breast.

Mothers feel the let-down reflex. Some describe it as a kind of tingling or drawing. After they have let down their milk, the breast feels heavy and ready for sucking. There is a real physical urge in the mother to apply the baby to the breast.

The psychological importance of the let-down reflex is that it is not completely automatic. It can be inhibited and conditioned. Conditioning takes place very easily, particularly during the first few weeks after birth. One mother I know delivered her baby at a hospital that always served the mother a cool glass of milk before nursing times. When she got home, she discovered her let-down reflex had been conditioned. When she saw a glass of milk, she started dripping milk. Many mothers get conditioned to their babies' hunger noises. They let down their milk at the first little noise their baby makes. Mothers away from their babies find they start

---

From *Marriage and Family Living*, 1958, *20*: 18–20. Reprinted with permission of the National Council on Family Relations.

dripping milk when they think of their babies about the time of a regular breast feeding. The time and the thought of the baby set off the conditioned reflex. That is why breast-fed babies may continue to get night feedings longer than bottle-fed babies. The problem is not only for the baby to get out of the habit of waking, but also for the mother to get out of the habit of letting down her milk in the middle of the night.

The let-down reflex can also be inhibited. Ely and Peterson (1941) demonstrated this with a brilliant series of experiments on cows. They banged paper bags in the cows' ears, and put cats on their backs. The cows stopped letting down their milk. Then oxytocin was injected and set off the reflex artificially. The Newtons demonstrated this same mechanism in a human being (1948). They studied a mother and baby in great detail over a period of several weeks. They found that each morning the baby got a predictable amount of milk from the breast at the first feeding. Then they distracted the mother in such a way as not to interfere with the baby. They put her feet in ice water, gave her electric shocks if she did not answer arithmetic questions correctly, and pulled her big toes. The baby sucked as usual, but got significantly less milk. On other days they distracted the mother, but injected oxytocin to set off the let-down artificially. On those days, in spite of the distraction, her let-down reflex worked artificially, and the baby got its usual milk.

It took pretty severe measures to inhibit the let-down reflex in the Newtons' experimental mother because her reflex was well conditioned. It takes much less to inhibit let-down in the tense new mother in the hospital. The Newtons later studied the symptoms of let-down in 103 mothers in a hospital (1950). Mothers who ended up with enough milk for their babies so that bottle supplementation was unnecessary had significantly more symptoms of let-down.

Even more important, mothers who went on to have plenty of milk for their babies were able to empty their breasts more completely through natural let-down. Mothers who did not give their babies enough milk were able to make only about half of their milk available before injections of oxytocin set off the let-down reflex artificially (Newton & Newton, 1950).

The first influence of the let-down reflex on the mother-child relationship is the fact that the child's supply of milk can only be secure and abundant when the let-down reflex is working well. It is a mistake to consider all breast feeding as a comforting, close experience. When the let-down reflex is inhibited, it can be an extremely frustrating experience. The baby is hungry. He has learned that food comes from the breast. He sucks the breast—but this time his mother is upset. Her usually easy and abundant supply is greatly decreased. The baby is frustrated and may fuss with rage.

The second way the let-down reflex can influence the mother-child relationship is that it is one of the physiological mechanisms contributing to the sensuous pleasure the mother gets from breast feeding. Dr. Spence (1938), the wise British pediatrician, so aptly described the look of "lascivious content" that comes over the faces of relaxed breast feeding mothers. Actually, what happens is that the let-down reflex acts not only on the breast, but also on the uterus; Dr. Chasser Moir (1934) made actual tracings of uterine contractions during breast feeding. The uterus not only contracted during breast feeding, but continued to contract rhythmically for as long as twenty minutes after breast feeding. Uterine contractions have also been demonstrated to occur in women during sexual excitement (Kinsey *et al.*, 1953).

The mother with the well-conditioned and uninhibited let-down reflex not only gives her baby the pleasure of a secure and abundant milk supply, but she, herself, derives pleasurable physical sensations from the act of breast feeding. She wants her baby, physically as well as mentally, and thus the mother-child relationship, like the husband-wife relationship, starts on the basis of physical desire as well as companionship.

Our own great-grandmothers breast-fed their children over a year, and people in very rural regions in the U.S.A. continue to do so today. Breast feeding lasts for more than two years in many other parts of the world. For instance, in Pakistan a survey of children eighteen months to twenty-four months of age showed that 100 per cent of them were still getting some breast feeding (Jelliffe, 1955). Ninety-two per cent were also receiving solid foods, but the point is that every one of them still had this close physical relationship with their mothers. Pakistan is not a bit unusual but quite typical of countries in which breast feeding is an accepted part of the cultural pattern.

A child who is breast-fed for this normal physiological period of two years has at least two thousand feedings. In a relaxed woman almost every one of the two thousand feedings gives her pleasurable sensations as well as satisfying the baby (Newton, 1955). This type of shared experience certainly is one that should not be overlooked when we think in terms of the mother-child relationship and ways of making it more constructive.

REFERENCES

ELY, F., & PETERSEN, W. E. Factors involved in the ejection of milk. *J. dairy Sci.*, 1941, *24*, 221 ff.

JELLIFFE, D. B. *Infant nutrition in the subtropics and tropics*. Geneva: World Health Organization, 1955.

KINSEY, A. C., *et al. Sexual behavior in the human female.* Philadelphia: W. B. Saunders Company, 1953.

MOIR, CHASSAR. Recording the contractions of the human pregnant and nonpregnant uterus. *Transactions Edinburgh Obstetrical Society,* 1934, *54,* 93 ff.

NEWTON, M., & NEWTON, N. The let-down reflex in human lactation. *J. Pediatrics,* 1948, *33,* 698 ff.

NEWTON, N. *Maternal emotions*. New York: Paul B. Hoeber, 1955.

NEWTON, N., & NEWTON, M. Relation of the let-down reflex to ability to breast feed. *Pediatrics,* 1950, *5,* 726–733.

PETERSEN, W. E. Effects of certain hormones and drugs on the perfused mammary gland. *Proceedings of the Society of Experimental Biology and Medicine,* 1942, *50,* 298 ff.

PETERSEN, W. E., & LUDWICK, T. M. The humoral nature of the factor causing the let down of milk. *Federation Proceedings,* 1942, *1,* 66 ff.

SPENCE, J. C. The modern decline of breast-feeding. *Brit. med. J.*, 1938, *2,* 729 ff.

17

Trends in Child Care over Three Generations

Leone Kell and Joan Aldous

In their research Kell and Aldous attack the child-rearing problem by looking at the attitudes of three generations—college students, their mothers, and their grandmothers—to see if views on child rearing have changed. Their aim is to test different theories of how children take over child-rearing values from their mothers rather than to explore the infant care situation itself. Their conclusion that these values have changed stems from the study of interview data, a method in contrast with the observational and experimental procedures utilized by other authors in Part II.

In the rapidly changing milieu of today do children tend to accept, modify, or reject their parents' child-rearing procedures and ideologies? What do they, in turn, transmit to their children? Are the trends differentiated by sex of child and by parent? Two contrasting theories are relevant. Parsons and Bales (1955) and Sears (1957) hold that children acquire their most fundamental values and attitudes from their parents, especially their mothers. It might therefore be assumed that children would accept and repeat their parents' patterns. On the other hand Mead (1955) suggests that in American culture children expect to behave differently from their parents and will likewise treat their children differently.

Interest in such questions was stimulated by exploratory interviews with fifty college students and their mothers. Answers to questions in the three areas of rules for behavior, discipline, and affection-giving were obtained from mothers in regard to their own mothers' patterns, and from the mothers' children who were students (twenty-seven girls, twenty-three boys), concerning their mother's ideas. Inquiry was made as to the students' plans for rearing their own children. Three generations were thus involved. Details of this exploratory study have been reported elsewhere (Kell & Aldous, 1960). Chi-square analysis was used on the data.

Comparisons of the mothers' child-rearing values with their mothers (the students' grandmothers) showed that only in the area of discipline were the two generations significantly (.001 level) more likely to have values alike than unlike. In the areas of rules of behavior and affection-giving there was no significant difference between the number of individuals whose values were like those of their mothers and the number of individuals whose values were unlike those of their mothers. The internalization theory was thus confirmed in one area, while neither it nor the discontinuity theory was upheld in the other two areas.

From *Marriage and Family Living,* 1960, *22,* 176–177. Reprinted with permission of the National Council on Family Relations.

In regard to the students a significantly greater number (.001 level) had values like their mothers' rather than unlike them, in the area of rules of behavior. The finding in the area of affection-giving was nonsignificant, as in the previous comparison of the mothers' generation with that of the students' grandmothers. In the area of discipline, however, there was a sex difference. The male students significantly more often possessed child-rearing values like those of their mothers, while this was not true of the girls.[1] The discontinuity theory, therefore, even when broadly interpreted was not upheld by any of the findings. The data, however, did uphold the internalization theory two times out of three for the male students and once out of three tests for the female students.

But what of the individuals who had some child-rearing values that did differ from those of their mothers? It was thought that these values of the previous generation would be replaced by values of the more permissive, developmental ideology advocated over the last 20 years by most child guidance authorities. These include unconditional affection-giving including physical demonstrations of affection; reasoning and talking to in the area of discipline, and in the area of behavioral standards, getting along with others, and encouraging the child to follow his own ideas and to develop his abilities at his own chosen rate.

The differences between the students' mothers and their grandmothers were in the expected direction. More of the mothers differed from their mothers by having permissive values than differed by having traditional values. Only in the area of rules of behavior, however, was the difference significant. Analysis of the child-rearing values of the students whose values were different from those of their mothers showed an interesting sex difference. The girls were significantly more likely to have permissive than traditional values in the discipline area, (.01 level) and in the area of rules of behavior and affection-giving the data came close (.06 level) to supporting the hypothesis. The hypothesis, however, was not confirmed in the case of the boys, and the findings in the area of rules of behavior, though not significant, were contrary to those hypothesized. The boys were more likely to differ from their mothers in approving traditional middle class rules of behavior for children than in having permissive standards.

The tests of the discontinuity and internalization theories suggested that the mothers and the female students had child-rearing values differing from their mothers oftener than did the male students. Thus, only in the area of discipline for the mothers and the area of rules of behavior for the female students, did the data from these two groups support the theory that children adopt their mother's child-rearing values. For the male students, the internalization theory was upheld by the data in the two areas of rules of behavior and discipline. The data from neither the mothers nor the students supported the discontinuity theory.

There was also a consistent trend as theorized in all areas for the mothers and female students who possessed child-rearing values unlike their mothers to differ in having permissive values. This trend among the mothers coupled with the number of them having values unlike their own mothers pointed to a fairly widespread change in child-rearing values that had occurred between the time the students' grandmothers were bringing up the students' mothers, and the time the students' mothers were rearing them.

REFERENCES

KELL, LEONE, & ALDOUS, JOAN. The relation between mothers' child rearing ideologies and their children's perceptions of maternal control. *Child Develpm.*, 1960, *31*, 145–156.

MEAD, MARGARET. Theoretical setting—1954. In Margaret Mead & Martha Wolfenstein (Eds.), *Childhood in contemporary cultures*. Chicago: University of Chicago Press, 1955. P. 11.

PARSONS, T., & BALES, R. F. *Family, socialization, and interaction processes*. Glencoe, Ill.: Free Press, 1955.

SEARS, R. R. Identification as a form of behavioral development. In D. R. Harris (Ed.), *The concept of development*. Minneapolis: University of Minnesota Press, 1957. Pp. 149–161.

---

[1] Previous chi-square analysis had shown an association significant at the .05 level between male sex and similarity of child-rearing values to those of their mothers in the area of discipline. For this reason data from the two sexes was tested separately in the further statistical analysis reported above. No significant association between sex and similarity of child-rearing values was found in the other two areas even at the comparatively low level of 10.

18

The Nature of Love

Harry F. Harlow

Perhaps some of the most significant research on the mother-infant relationship is that of Harlow at the University of Wisconsin. This paper and the next constitute reports on his findings. The use of monkeys permits experimentation not ethically possible with human infants, but the experimental situations can be associated with similar situations in the human world. The results indicate that other variables in mothering may be more significant than feeding. The reader should study this paper in relation to Newton's (16).

Love is a wondrous state, deep, tender, and rewarding. Because of its intimate and personal nature it is regarded by some as an improper topic for experimental research. But, whatever our personal feelings may be, our assigned mission as psychologists is to analyze all facets of human and animal behavior into their component variables. So far as love or affection is concerned, psychologists have failed in this mission. The little we know about love does not transcend simple observation, and the little we write about it has been written better by poets and novelists. But of greater concern is the fact that psychologists tend to give progressively less attention to a motive which pervades our entire lives. Psychologists, at least psychologists who write textbooks, not only show no interest in the origin and development of love or affection, but they seem to be unaware of its very existence.

The apparent repression of love by modern psychologists stands in sharp contrast with the attitude taken by many famous and normal people. The word "love" has the highest reference frequency of any word cited in Bartlett's book of *Familiar Quotations*. It would appear that this emotion has long had a vast interest and fascination for human beings, regardless of the attitude taken by psychologists; but the quotations cited, even by famous and normal people, have a mundane redundancy. These authors and authorities have stolen love from the child and infant and made it the exclusive property of the adolescent and adult.

Thoughtful men, and probably all women, have speculated on the nature of love. From the developmental point of view, the general plan is quite clear: The initial love responses of the human being are those made by the infant to the mother or some mother surrogate. From this intimate attachment of the child to the mother, multiple learned and generalized affectional responses are formed.

Unfortunately, beyond these simple facts we know little about the fundamental variables underlying the formation of affectional responses and little about the mechanisms through which the love of the infant for the mother develops into the multifaceted response patterns characterizing love or affection in the adult. Because of the dearth of experimentation, theories about the fundamental nature of affection have evolved at the level of observation, intuition, and discerning guesswork, whether these have been proposed by psychologists, sociologists, anthropologists, physicians, or psychoanalysts.

The position commonly held by psychologists and sociologists is quite clear: The basic motives are, for the most part, the primary drives—particularly hunger, thirst, elimination, pain, and sex—and all other motives, including love or affection, are derived or secondary drives. The mother is associated with the reduction of the primary drives—particularly hunger, thirst, and pain—and through learning, affection or love is derived.

It is entirely reasonable to believe that the mother through association with food may become a secondary-reinforcing agent, but this is an inadequate mechanism to account for the persistence of the infant-maternal ties. There is a spate of researches on the formation of secondary reinforcers to hunger and thirst reduction. There can be no question that almost any external stimulus can become a secondary reinforcer if properly associated with tissue-need reduction, but the fact remains that this redundant literature demonstrates unequivocally that such derived drives suffer relatively rapid experimental extinction. Contrariwise, human affection does not extinguish when the mother ceases to have intimate association with the drives in question. Instead, the affectional ties to the mother show a lifelong, unrelenting persistence and, even more surprising, widely expanding generality.

Oddly enough, one of the few psychologists who took a position counter to modern psychological dogma was John B. Watson, who believed that love was an innate emotion elicited by cutaneous stimulation of the erogenous zones. But

---

From *American Psychologist*, 1958, *13*: 673–685.

experimental psychologists, with their peculiar propensity to discover facts that are not true, brushed this theory aside by demonstrating that the human neonate had no differentiable emotions, and they established a fundamental psychological law that prophets are without honor in their own profession.

The psychoanalysts have concerned themselves with the problem of the nature of the development of love in the neonate and infant, using ill and aging human beings as subjects. They have discovered the overwhelming importance of the breast and related this to the oral erotic tendencies developed at an age preceding their subjects' memories. Their theories range from a belief that the infant has an innate need to achieve and suckle at the breast to beliefs not unlike commonly accepted psychological theories. There are exceptions, as seen in the recent writings of John Bowlby, who attributes importance not only to food and thirst satisfaction, but also to "primary object-clinging," a need for intimate physical contact, which is initially associated with the mother.

As far as I know, there exists no direct experimental analysis of the relative importance of the stimulus variables determining the affectional or love responses in the neonatal and infant primate. Unfortunately, the human neonate is a limited experimental subject for such researches because of his inadequate motor capabilities. By the time the human infant's motor responses can be precisely measured, the antecedent determining conditions cannot be defined, having been lost in a jumble and jungle of confounded variables.

Many of these difficulties can be resolved by the use of the neonatal and infant macaque monkey as the subject for the analysis of basic affectional variables. It is possible to make precise measurements in this primate beginning at two to ten days of age, depending upon the maturational status of the individual animal at birth. The macaque infant differs from the human infant in that the monkey is more mature at birth and grows more rapidly; but the basic responses relating to affection, including nursing, contact, clinging, and even visual and auditory exploration, exhibit no fundamental differences in the two species. Even the development of perception, fear, frustration, and learning capability follows very similar sequences in rhesus monkeys and human children.

Three years' experimentation before we started our studies on affection gave us experience with the neonatal monkey. We had separated more than 60 of these animals from their mothers 6 to 12 hours after birth and suckled them on tiny bottles. The infant mortality was only a small fraction of what would have obtained had we let the monkey mothers raise their infants. Our bottle-fed babies were healthier and heavier than monkey-mother-reared infants. We know that we are better monkey mothers than are real monkey mothers thanks to synthetic diets, vitamins, iron extracts, penicillin, chloromycetin, 5% glucose, and constant, tender, loving care.

During the course of these studies we noticed that the laboratory-raised babies showed strong attachment to the cloth pads (folded gauze diapers) which were used to cover the hardware-cloth floors of their cages. The infants clung to

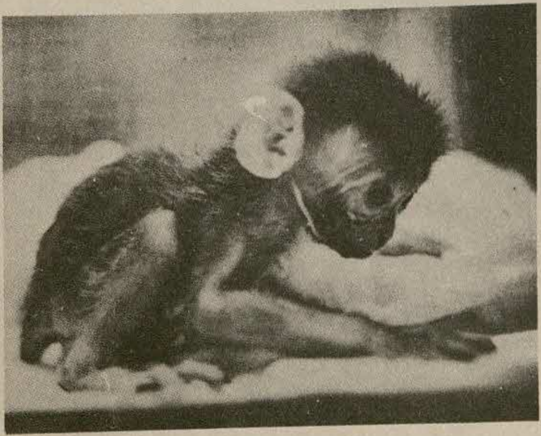


Fig. 1: *Response to cloth pad by one-day-old monkey..*

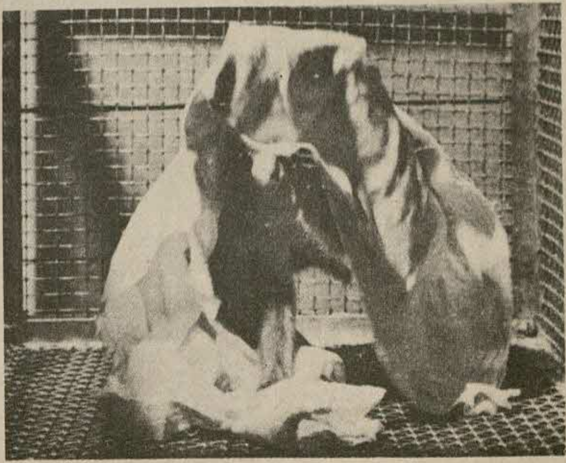


Fig. 2: *Response to gauze pad by six-month-old monkey used in earlier study.*

these pads and engaged in violent temper tantrums when the pads were removed and replaced for sanitary reasons. Such contact-need or responsiveness had been reported previously by Gertrude van Wagenen for the monkey and by Thomas McCulloch and George Haslerud for the chimpanzee and is reminiscent of the devotion often exhibited by human infants to their pillows, blankets, and soft, cuddly stuffed toys. Responsiveness by the one-day-old infant monkey to the cloth pad is shown in Figure 1, and an unusual and strong attachment of a six-month-old infant to the cloth pad is illustrated in Figure 2. The baby, human or monkey, if it is to survive, must clutch at more than a straw.

We had also discovered during some allied observational studies that a baby monkey raised on a bare wire-mesh cage floor survives with difficulty, if at all, during the first five days of life. If a wire-mesh cone is introduced, the baby does better; and, if the cone is covered with terry cloth, husky, healthy, happy babies evolve. It takes more than a baby and a box to make a normal monkey. We were impressed by the possibility that, above and beyond the bubbling fountain of breast or bottle, contact comfort might be a very important variable in the development of the infant's affection for the mother.

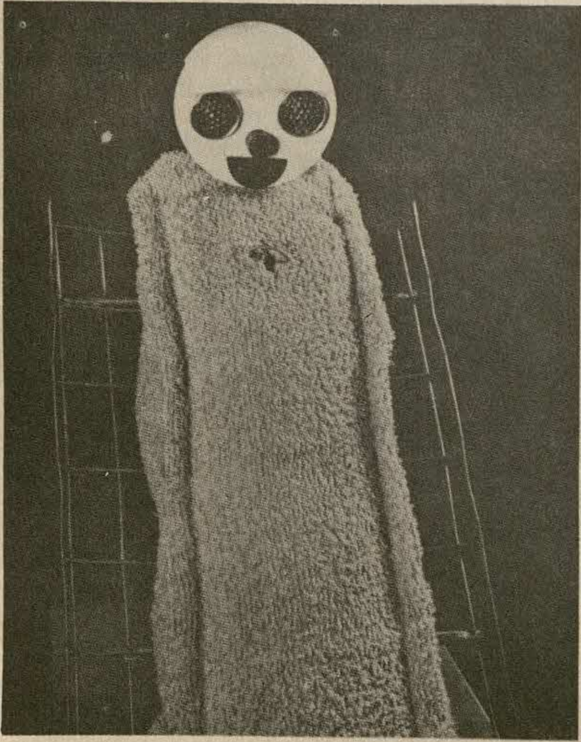


Fig. 3: *Cloth mother surrogate.*

At this point we decided to study the development of affectional responses of neonatal and infant monkeys to an artificial, inanimate mother, and so we built a surrogate mother which we hoped and believed would be a good surrogate mother. In devising this surrogate mother we were dependent neither upon the capriciousness of evolutionary processes nor upon mutations produced by chance radioactive fallout. Instead, we designed the mother surrogate in terms of modern human-engineering principles (Figure 3). We produced a perfectly proportioned, streamlined body stripped of unnecessary bulges and appendices. Redundancy in the surrogate mother's system was avoided by reducing the number of breasts from two to one and placing this unibreast in an upper-thoracic, sagittal position, thus maximizing the natural and known perceptual-motor capabilities of the infant operator. The surrogate was made from a block of wood, covered with sponge rubber, and sheathed in tan cotton terry cloth. A light bulb behind her radiated heat. The result was a mother, soft, warm, and tender, a mother with infinite patience, a mother available twenty-four hours a day, a mother that never scolded her infant and never struck or bit her baby in anger. Furthermore, we designed a mother-machine with maximal maintenance efficiency since failure of any system or function could be resolved by the simple substitution of black boxes and new component parts. It is our opinion that we engineered a very superior monkey mother, al-

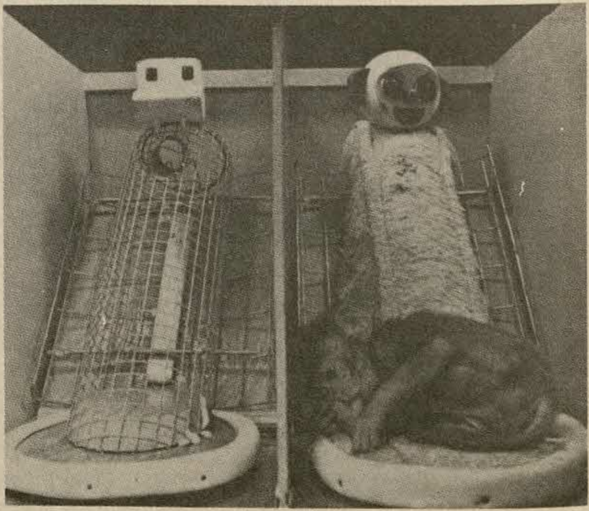


Fig. 4: *Wire and cloth mother surrogates.*

though this position is not held universally by the monkey fathers.

Before beginning our initial experiment we also designed and constructed a second mother surrogate, a surrogate in which we deliberately built less than the maximal capability for contact comfort. This surrogate mother is illustrated in Figure 4. She is made of wire-mesh, a substance entirely adequate to provide postural support and nursing capability, and she is warmed by radiant heat. Her body differs in no essential way from that of the cloth mother surrogate other than in the quality of the contact comfort which she can supply.

In our initial experiment, the dual mother-surrogate condition, a cloth mother and a wire mother were placed in different cubicles attached to the infant's living cage as shown in Figure 4. For four newborn monkeys the cloth mother lactated and the wire mother did not; and, for the other four, this condition was reversed. In either condition the infant received all its milk through the mother surrogate as soon as it was able to maintain itself in this way, a capability achieved within two or three days except in the case of very immature infants. Supplementary feedings were given until the milk intake from the mother surrogate was adequate. Thus, the experiment was designed as a test of the relative importance of the variables of contact comfort and nursing comfort. During the first 14 days of life the monkey's cage floor was covered with a heating pad wrapped in a folded gauze diaper, and thereafter the cage floor was bare. The infants were always free to leave the heating pad or cage floor to contact either mother, and the time spent on the surrogate mothers was automatically recorded. Figure 5 shows the total time spent on the cloth and wire mothers under the two conditions of feeding. These data make it obvious that contact comfort is a variable of overwhelming importance in the development of affectional responses, whereas lactation is a variable of negligible importance. With age and opportunity to learn, subjects with the lactating wire mother showed decreasing responsiveness to her and increasing responsiveness to the non-lactating cloth mother, a finding completely contrary to any interpretation of derived drive in which the mother-form becomes conditioned to hunger-thirst reduction. The persistence of these differential responses throughout 165 consecutive days of testing is evident in Figure 6.

One control group of neonatal monkeys was raised on a single wire mother, and a second control group was raised on a single cloth mother. There were no differences between these two groups in amount of milk ingested or in weight gain. The only difference between the groups lay in the composition of the feces, the softer stools of the wire-mother infants suggesting psy-

Fig. 5: *Time spent on cloth and wire mother surrogates.*

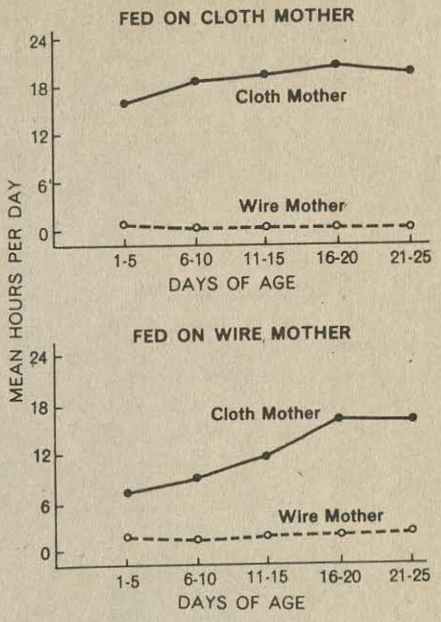
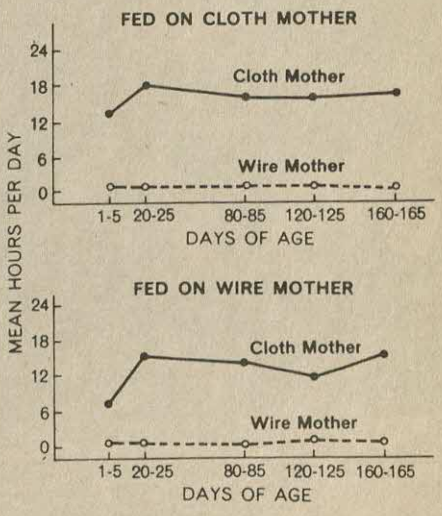


Fig. 6: *Long-term contact time on cloth and wire mother surrogates.*



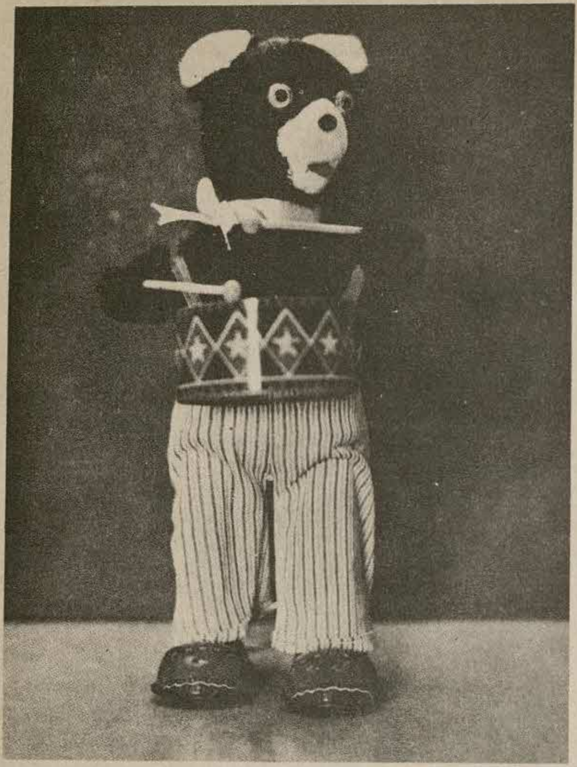


Fig. 7: *Typical fear stimulus.*

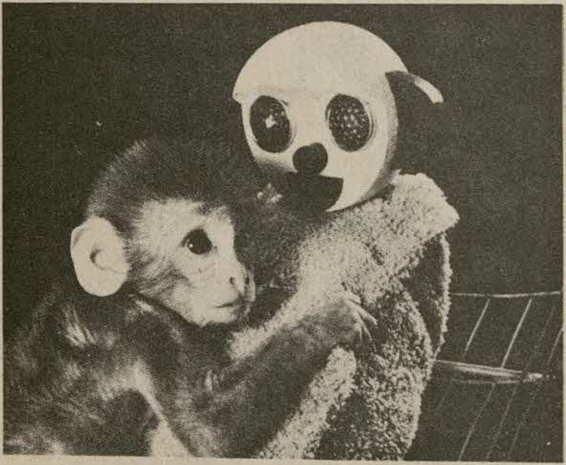


Fig. 8: *Typical response to cloth mother surrogate in fear test.*

chosomatic involvement. The wire mother is biologically adequate but psychologically inept.

We were not surprised to discover that contact comfort was an important basic affectional or love variable, but we did not expect it to overshadow so completely the variable of nursing; indeed, the disparity is so great as to suggest that the primary function of nursing as an affectional variable is that of insuring frequent and intimate body contact of the infant with the mother. Certainly, man cannot live by milk alone. Love is an emotion that does not need to be bottle- or spoon-fed, and we may be sure that there is nothing to be gained by giving lip service to love.

A charming lady once heard me describe these experiments; and, when I subsequently talked to her, her face brightened with sudden insight: "Now I know what's wrong with me," she said, "I'm just a wire mother." Perhaps she was lucky. She might have been a wire wife. . . .

One function of the real mother, human or subhuman, and presumably of a mother surrogate, is to provide a haven of safety for the infant in times of fear and danger. The frightened or ailing child clings to its mother, not its father; and this selective responsiveness in times of distress, disturbance, or danger may be used as a measure of the strength of affectional bonds. We have tested this kind of differential responsiveness by presenting to the infants in their cages, in the presence of the two mothers, various fear-producing stimuli such as the moving toy bear illustrated in Figure 7. A typical response to a fear stimulus is shown in Figure 8, and the data on differential responsiveness are presented in Figure 9. It is apparent that the cloth mother is highly preferred over the wire one, and this differential selectivity is enhanced by age and experience. In this situation, the variable of nursing appears to be of absolutely no importance: the infant consistently seeks the soft mother surrogate regardless of nursing condition.

Similarly, the mother or mother surrogate provides its young with a source of security, and this role or function is seen with special clarity when mother and child are in a strange situation. At the present time we have completed tests for this relationship on four of our eight baby monkeys assigned to the dual mother-surrogate condition by introducing them for three minutes into the strange environment of a room measuring six feet by six feet by six feet (also called the "open-field test") and containing multiple stimuli known to elicit curiosity-manipulatory responses in baby monkeys. The subjects were placed in this situation twice a week for eight weeks with no mother surrogate present during alternate sessions and the cloth mother present during the others. A cloth diaper was always available as one of the stimuli throughout all sessions. After one or two adaptation sessions, the infants always rushed to the mother surrogate when she was present and clutched her, rubbed their bodies against her, and frequently manipulated her body and face. After a few additional sessions, the infants began to use the mother surrogate as a source of security, a base of operations. They would explore and manipulate a stimulus and then return to the mother before adventuring again into the strange new world. The behavior of these infants was quite different when the mother was absent from the room. Frequently they would freeze in a crouched position. Emotionality indices such as vocalization, crouching, rocking, and sucking increased sharply, as shown in Figure 10. Total emotionality score was cut in half when the mother was present. In the absence of the mother some of the experimental monkeys would rush to the center of the room where the mother was customarily placed and then run rapidly from object to object, screaming and crying all the while. Continuous, frantic clutching of their bodies was very common, even when not in the crouching position. These monkeys frequently contacted and clutched the cloth diaper, but this action never pacified them. The same behavior occurred in the presence of the wire mother. No difference between the cloth-mother-fed and wire-mother-fed infants was demonstrated under either condition. Four control infants never raised with a mother surrogate showed the same emotionality scores when the mother was absent as the experimental infants showed in the absence of the mother, but the controls' scores were slightly larger in the presence of the mother surrogate than in her absence.

Fig. 9: *Differential responsiveness in fear tests.*

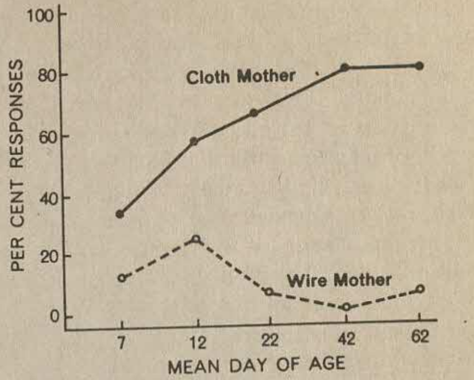
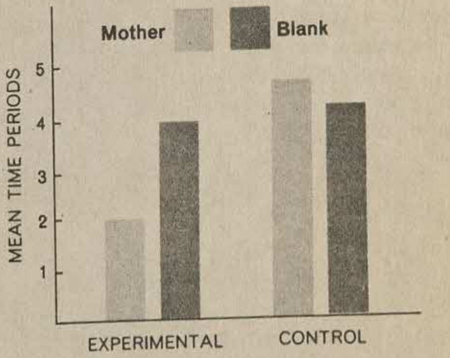


Fig. 10: *Emotionality index with and without the presence of the cloth mother.*



Some years ago Robert Butler demonstrated that mature monkeys enclosed in a dimly lighted box would open and reopen a door hour after hour for no other reward than that of looking outside the box. We now have data indicating that neonatal monkeys show this same compulsive visual curiosity on their first test day in an adaptation of the Butler apparatus which we call the "love machine," an apparatus designed to measure love. Usually these tests are begun when the monkey is 10 days of age, but this same persistent visual exploration has been obtained in a three-day-old monkey during the first half-hour of testing. Butler also demonstrated that rhesus monkeys show selectivity in rate and frequency of door-opening to stimuli of differential attractiveness in the visual field outside the box. We have utilized this principle of response selectivity by the monkey to measure strength of affectional responsiveness in our infants in the baby version of the Butler box. The test sequence involves four repetitions of a test battery in which four stimuli—cloth mother, wire mother, infant monkey, and empty box—are presented for a 30-minute period on successive days. The first four subjects in the dual mother-surrogate group were given a single test sequence at 40 to 50 days of age, depending upon the availability of the apparatus, and only their data are presented. The second set of four subjects is being given repetitive tests to obtain information relating to the development of visual exploration. The apparatus is illustrated in Figure 11. The data obtained from the first four infants raised with the two mother surrogates are presented in the middle graph of Figure 12 and show approximately equal responding to the cloth mother and another infant monkey, and no greater responsiveness to the wire mother than to an empty box. Again, the results are independent of the kind of mother that lactated, cloth or wire. The same results are found for a control group raised, but not fed, on a single cloth mother; these data appear in the graph on the right. Contrariwise, the graph on the left shows no differential responsiveness to cloth and wire mothers by a second control group, which was not raised on any mother surrogate. We can be certain that not all love is blind.

The first four infant monkeys in the dual mother-surrogate group were separated from their mothers between 165 and 170 days of age and tested for retention during the following 9 days and then at 30-day intervals for six successive months. Affectional retention as measured by the modified Butler box is given in Figure 13. In keeping with the data obtained on adult monkeys by Butler, we find a high rate of responding to any stimulus, even the empty box. But throughout the entire 185-day retention period there is a consistent and significant difference in response frequency to the cloth mother contrasted with either the wire mother or the empty box, and no consistent difference between wire mother and empty box.

Affectional retention was also tested in the open field during the first 9 days after separation and then at 30-day intervals, and each test condition was run twice at each retention interval. The infant's behavior differed from that observed during the period preceding separation. When the cloth mother was present in the post-separation period, the babies rushed to her, climbed up, clung tightly to her, and rubbed their heads and faces against her body. After this initial embrace and reunion, they played on the mother, including biting and tearing at her cloth cover; but they rarely made any attempt to leave her during the test period, nor did they manipulate or play with the objects in the room, in contrast with their behavior before maternal separation. The only exception was the occasional monkey that left the mother surrogate momentarily, grasped the folded piece of paper (one of the standard stimuli in the field), and brought it quickly back to the mother. It appeared that deprivation had enhanced the tie to the mother and rendered the contact-comfort need so prepotent that need for the mother overwhelmed the exploratory motives during the brief, three-minute test sessions. No change in these behaviors was observed throughout the 185-day period. When the mother was absent from the open field, the behavior of the infants was similar in the initial retention test to that during the preseparation tests; but they tended to show gradual adaptation to the open-field situation with repeated testing and, consequently, a reduction in their emotionality scores.

In the last five retention test periods, an additional test was introduced in which the surrogate mother was placed in the center of the room and covered with a clear Plexiglas box. The monkeys were initially disturbed and frustrated when their explorations and manipulations of the box failed to provide contact with the mother. However, all animals adapted to the situation rather rapidly. Soon they used the box as a place of orientation for exploratory and play behavior, made frequent contacts with the objects in the field, and very often brought these objects to the Plexiglas box. The emotionality index was slightly higher than in the condition of the available cloth mothers, but it in no way approached the emotionality level displayed when the cloth mother was absent. Obviously, the infant mon-

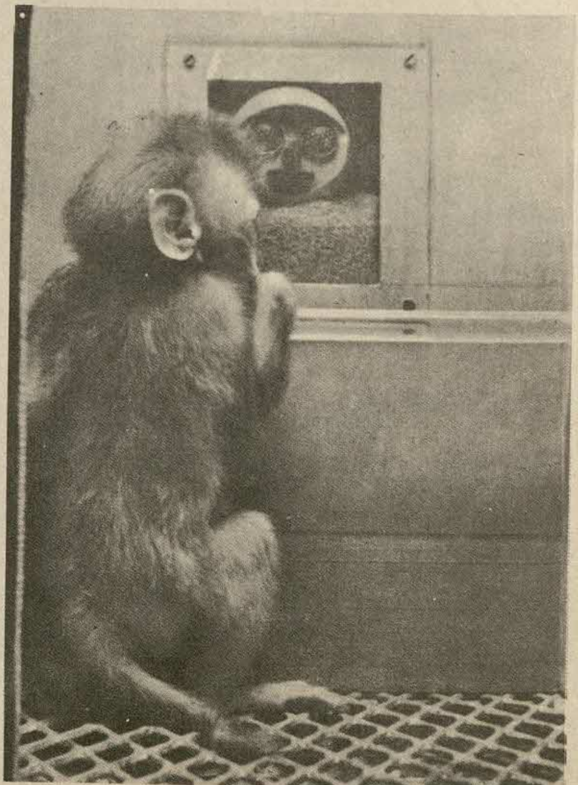


Fig. 11: *Visual exploration apparatus.*

keys gained emotional security by the presence of the mother even though contact was denied.

Affectional retention has also been measured by tests in which the monkey must unfasten a three-device mechanical puzzle to obtain entrance into a compartment containing the mother surrogate. All the trials are initiated by allowing the infant to go through an unlocked door, and in half the trials it finds the mother present and in half, an empty compartment. The door is then locked and a ten-minute test conducted. In tests given prior to separation from the surrogate mothers, some of the infants had solved this puzzle and others had failed. The data of Figure 14 show that on the last test before separation there were no differences in total manipulation under mother-present and mother-absent conditions, but striking differences exist between the two conditions throughout the post-separation test periods. Again, there is no interaction with conditions of feeding.

The over-all picture obtained from surveying the retention data is unequivocal. There is little, if any, waning of responsiveness to the mother throughout this five-month period as indicated by any measure. It becomes perfectly obvious that this affectional bond is highly resistant to forgetting and that it can be retained for very long periods of time by relatively infrequent contact reinforcement. During the next year, retention tests will be conducted at 90-day intervals, and further plans are dependent upon the results obtained. It would appear that affectional responses may show as much resistance to extinction as has been previously demonstrated for learned fears and learned pain, and such data would be in keeping with those of common human observation. . . .

We have already described the group of four control infants that had never lived in the presence of any mother surrogate and had demonstrated no sign of affection or security in the presence of the cloth mothers introduced in test sessions. When these infants reached the age of 250 days, cubicles containing both a cloth mother and a wire mother were attached to their cages. There was no lactation in these mothers, for the monkeys were on a solid-food diet. The initial reaction of the monkeys to the alterations was one of extreme disturbance. All the infants screamed violently and made repeated attempts to escape the cage whenever the door was opened. They kept a maximum distance from the mother surrogates and exhibited a considerable amount of rocking and crouching behavior, indicative of emotionality. Our first thought was

Fig. 12: *Differential responses to visual exploration.*

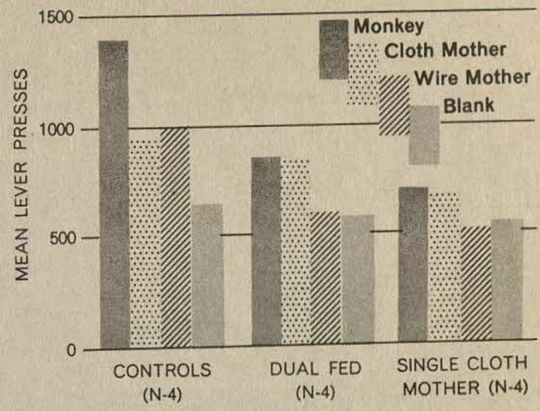


Fig. 13: *Retention of differential visual-exploration responses.*

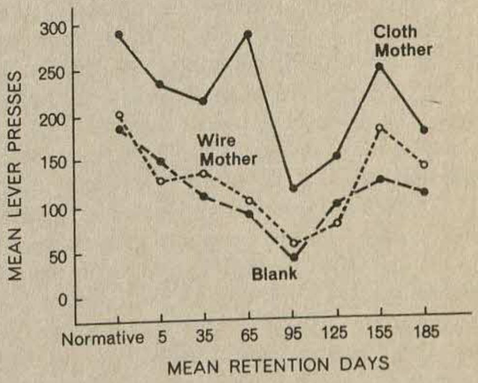


Fig. 14: *Retention of puzzle manipulation responsiveness.*

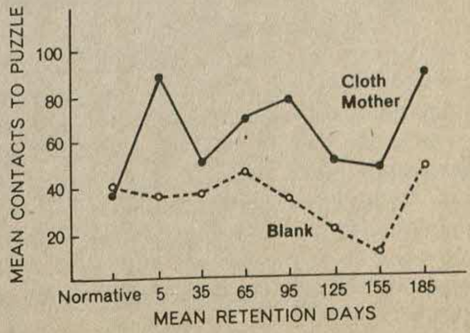
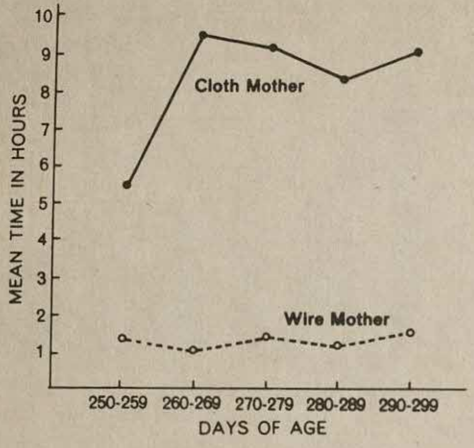


Fig. 15: *Differential time spent on cloth and wire mother surrogates by monkeys started at 250 days of age.*



that the critical period for the development of maternally directed affection had passed and that these macaque children were doomed to live as affectional orphans. Fortunately, these behaviors continued for only 12 to 48 hours and then gradually ebbed, changing from indifference to active contact on, and exploration of, the surrogates. The home-cage behavior of these control monkeys slowly became similar to that of the animals raised with the mother surrogates from birth. Their manipulation and play on the cloth mother became progressively more vigorous to the point of actual mutilation, particularly during the morning after the cloth mother had been given her daily change of terry covering. The control subjects were now actively running to the cloth mother when frightened and had to be coaxed from her to be taken from the cage for formal testing.

Objective evidence of these changing behaviors is given in Figure 15, which plots the amount of time these infants spent on the mother surrogates. Within 10 days mean contact time is approximately nine hours, and this measure remains relatively constant throughout the next 30 days. Consistent with the results on the subjects reared from birth with dual mothers, these late-adopted infants spent less than one and one-half hours per day in contact with the wire mothers, and this activity level was relatively constant throughout the test sessions. Although the maximum time that the control monkeys spent on the cloth mother was only about half that spent by the original dual mother-surrogate group, we cannot be sure that this discrepancy is a function of differential early experience. The control monkeys were about three months older when the mothers were attached to their cages than the experimental animals had been when their mothers were removed and the retention tests begun. Thus, we do not know what the amount of contact would be for a 250-day-old animal raised from birth with surrogate mothers. Nevertheless, the magnitude of the differences and the fact that the contact-time curves for the mothered-from-birth infants had remained constant for almost 150 days suggest that early experience with the mother is a variable of measurable importance.

The control group has also been tested for differential visual exploration after the introduction of the cloth and wire mothers; these behaviors are plotted in Figure 16. By the second test session a high level of exploratory behavior had developed, and the responsiveness to the wire mother and the empty box is significantly greater than that to the cloth mother. This is probably not an artifact since there is every reason to believe that the face of the cloth mother is a fear stimulus to most monkeys that have not had extensive experience with this object during the first 40 to 60 days of life. Within the third test session a sharp change in trend occurs, and the cloth mother is then more frequently viewed than the wire mother or the blank box; this trend continues during the fourth session, producing a significant preference for the cloth mother.

Before the introduction of the mother surrogate into the home-cage situation, only one of the four control monkeys had ever contacted the cloth mother in the open-field tests. In general, the surrogate mother not only gave the infants no security, but instead appeared to serve as a fear stimulus. The emotionality scores of these control subjects were slightly higher during the mother-present test sessions than during the mother-absent test sessions. These behaviors were changed radically by the fourth post-introduction test approximately 60 days later. In the absence of the cloth mothers the emotionality index in this fourth test remains near the earlier level, but the score is reduced by half when the mother is present, a result strikingly similar to that found for infants raised with the dual mother-surrogates from birth. The control infants now show increasing object exploration and play behavior, and they begin to use the mother as a base of operations, as did the infants raised from birth with the mother surrogates. However, there are still definite differences in the behavior of the two groups. The control infants do not rush directly to the mother and clutch her violently; but instead they go toward, and orient around, her, usually after an initial period during which they frequently show disturbed behavior, exploratory behavior, or both.

That the control monkeys develop affection or love for the cloth mother when she is introduced into the cage at 250 days of age cannot be questioned. There is every reason to believe, however, that this interval of delay depresses the intensity of the affectional response below that of the infant monkeys that were surrogate-mothered from birth onward. In interpreting these data it is well to remember that the control monkeys had had continuous opportunity to observe and hear other monkeys housed in adjacent cages and that they had had limited opportunity to view and contact surrogate mothers in the test situations, even though they did not exploit the opportunities.

During the last two years we have observed the behavior of two infants raised by their own mothers. Love for the real mother and love for the surrogate mother appear to be very similar. The baby macaque spends many hours a day clinging to its real mother. If away from the mother when frightened, it rushes to her and in her presence shows comfort and composure. As far as we can observe, the infant monkey's affection for the real mother is strong, but no stronger than that of the experimental monkey for the surrogate cloth mother, and the security that the infant gains from the presence of the real mother is no greater than the security it gains from a cloth surrogate. Next year we hope to put this problem to final, definitive, experimental test. But, whether the mother is real or a cloth surrogate, there does develop a deep and abiding bond between mother and child. In one case it may be the call of the wild and in the other the McCall of civilization, but in both cases there is "togetherness."

In spite of the importance of contact comfort, there is reason to believe that other variables of measurable importance will be discovered. Postural support may be such a variable, and it has been suggested that, when we build arms into the mother surrogate, 10 is the minimal number required to provide adequate child care. Rocking motion may be such a variable, and we are comparing rocking and stationary mother surrogates and inclined planes. The differential responsiveness to cloth mother and cloth-covered inclined plane suggests that clinging as well as contact is an affectional variable of importance. Sounds, particularly natural, maternal sounds, may operate as either unlearned or learned affectional variables. Visual responsiveness may be such a variable, and it is possible that some semblance of visual imprinting may develop in the neonatal monkey. There are indications that this becomes a variable of importance during the course of infancy through some maturational process.

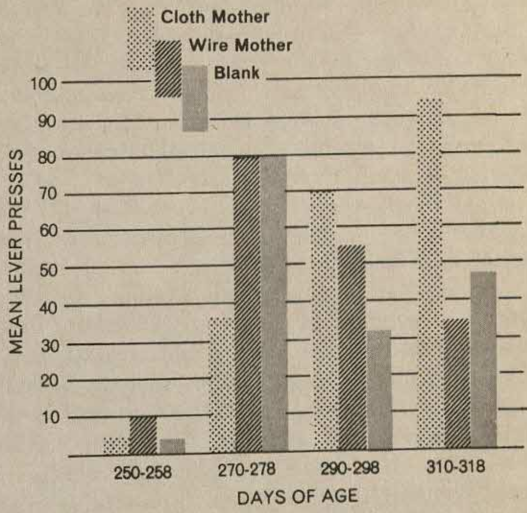


Fig. 16: *Differential visual exploration of monkeys started at 250 days of age.*

John Bowlby has suggested that there is an affectional variable which he calls "primary object following," characterized by visual and oral search of the mother's face. Our surrogate-mother-raised baby monkeys are at first inattentive to her face, as are human neonates to human mother faces. But by 30 days of age ever-increasing responsiveness to the mother's face appears—whether through learning, maturation, or both—and we have reason to believe that the face becomes an object of special attention.

Our first surrogate-mother-raised baby had a mother whose head was just a ball of wood since the baby was a month early and we had not had time to design a more esthetic head and face. This baby had contact with the blank-faced mother for 180 days and was then placed with two cloth mothers, one motionless and one rocking, both being endowed with painted, ornamented faces. To our surprise the animal would compulsively rotate both faces 180 degrees so that it viewed only a round, smooth face and never the painted, ornamented face. Furthermore, it would do this as long as the patience of the experimenter in reorienting the faces persisted. The monkey showed no sign of fear or anxiety, but it showed unlimited persistence. Subsequently it improved its technique, compulsively removing the heads and rolling them into its cage as fast as they were returned. We are intrigued by this observation, and we plan to examine systematically the role of the mother face in the development of infant-monkey affections. Indeed, these observations suggest the need for a series of ethological-type researches on the two-faced female.

Although we have made no attempts thus far to study the generalization of infant-macaque affection or love, the techniques which we have developed offer promise in this uncharted field. Beyond this, there are few if any technical difficulties in studying the affection of the actual, living mother for the child, and the techniques developed can be utilized and expanded for the analysis and developmental study of father-infant and infant-infant affection.

Since we can measure neonatal and infant affectional responses to mother surrogates, and since we know they are strong and persisting, we are in a position to assess the effects of feeding and contactual schedules; consistency and inconsistency in the mother surrogates; and early, intermediate, and late maternal deprivation. Again, we have here a family of problems of fundamental interest and theoretical importance.

If the researches completed and proposed make a contribution, I shall be grateful; but I have also given full thought to possible practical applications. The socioeconomic demands of the present and the threatened socioeconomic demands of the future have led the American woman to displace, or threaten to displace, the American man in science and industry. If this process continues, the problem of proper child-rearing practices faces us with startling clarity. It is cheering in view of this trend to realize that the American male is physically endowed with all the really essential equipment to compete with the American female on equal terms in one essential activity: the rearing of infants. We now know that women in the working classes are not needed in the home because of their primary mammalian capabilities; and it is possible that in the foreseeable future neonatal nursing will not be regarded as a necessity, but as a luxury—to use Veblen's term—a form of conspicuous consumption limited perhaps to the upper classes. But whatever course history may take, it is comforting to know that we are now in contact with the nature of love.

19

The Heterosexual Affectional System in Monkeys

Harry F. Harlow

This report presents a follow-up on the monkeys reared with surrogate mothers and illustrates clearly the vital importance of diadic relationships. Monkeys reared without monkey mothers had difficulty behaving like monkeys when they grew up and had their own offspring. This experiment stresses the role of experience in development and thus ties in with the work of Hunt (1) and Piaget (6).

The inspiration for this address came from observational data obtained from seven guinea pigs—two males and three females in a colony and two females brought in temporarily. Observations were provided by my ten-year-old daughter Pamela. These observations were made with love and endearment, and the behavior observed was endearment and love. Furthermore, these observations were made at a level of objectivity difficult for an adult to attain in this field.

Male and female guinea pigs are very fond of each other. They stare blissfully into the limpid pink or ruby or midnight-blue pools of each other's eyes. They nuzzle and they cuddle and the end production is not characterized by rush or rape. After all, one does not have to hurry if there is no hurry to be had. This, Pamela has witnessed several times. A caged, virgin adult female was brought by a friend for mating. Twirp, Pamela's large, black, gentle male, was put into the cage with the new female. He purred, nuzzled her, brushed up against her, smelled and licked her, and gradually conquered the frightened animal. A half-hour later they were snuggled up next to each other, peace-

---

From *American Psychologist*, 1962, *17*: 1–9

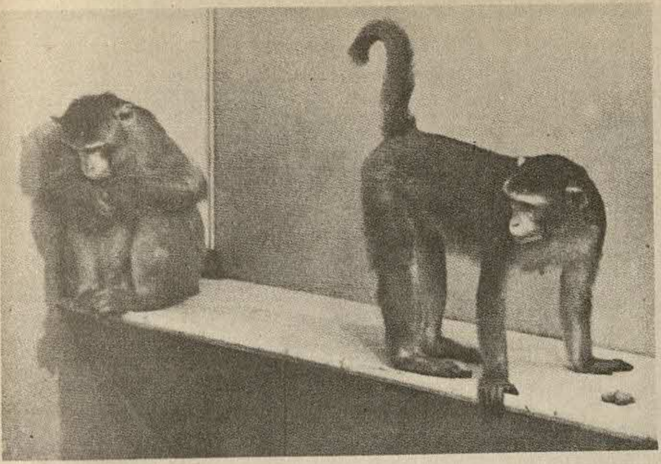


Fig. 1: *Initial response to female sexual-present posture. The male subsequently accepted the invitation.*

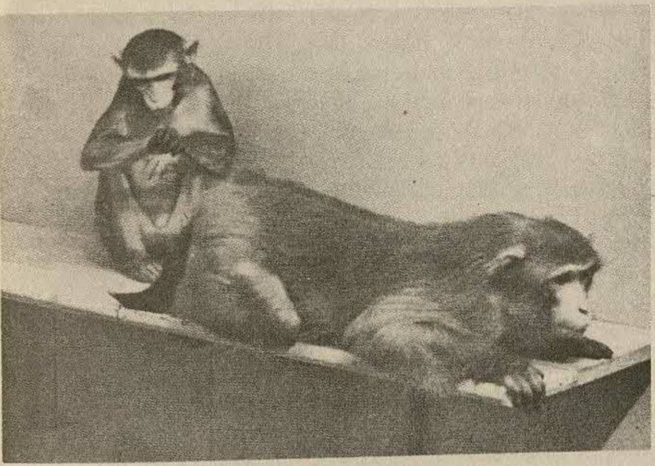


Fig. 2: *Initial response to male sexual-present posture. The female (No. 48) subsequently approached and groomed the male.*

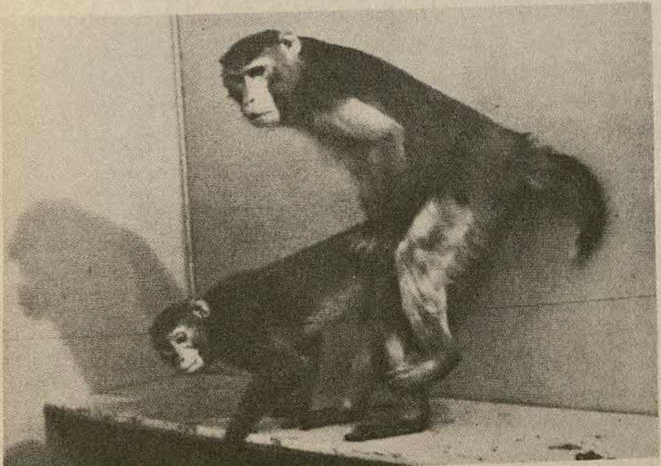


Fig. 3: *Normal male and female sexual positioning.*

ful and content, and they lived in bliss for several weeks until another friend brought in her female and Twirp repeated his patient, gentle approach. Twirp has convinced me that some male guinea pigs, at least, are endowed with an innate sense of decency, and I am happy to say that this is the way most male monkeys behave. I presume that there are some men who have as deep a depth of dignity as guinea pigs.

The guest stands, unfortunately, ended peaceful coexistence in the colony. For many months the five adult guinea pigs had lived amiably in one large cage, with Twirp in command and the second male playing second fiddle. While Twirp was host to the visiting females, White Patch commanded the permanent harem. When Twirp was reintroduced to the colony cage, it took but ten seconds to discover that he would not be tolerated. White Patch bared his teeth and lunged at Twirp, and to save the males, a new cage was acquired.

This led to various divisions of the females and led Pamela to discover particular male guinea pigs like particular female guinea pigs, and they squeal piteously when separated, even when the female is so bulging with babies that she can offer the male nothing in terms of drive reduction. Particular female guinea pigs like particular male guinea pigs. Tastes seem fairly stable, for even after weeks of peaceful residence with the unfavored male, the female will still attempt to get to her favorite male, and after weeks of quiet residence with unfavored females, the male will still try to get to his favorite female.

The females, like the males, defend their rights. In the happy one-cage days two females were separated from the group to care for their litters. White Thrush, in an advanced stage of pregnancy, lived alone with the males. When Chirp was returned to the colony cage after three weeks of maternal chores, both males approached enthusiastically, making friendly gestures. But Hell hath no fury like a female guinea pig spurned, and White Thrush would not tolerate infidelity. She hissed at Chirp, and lunged, and as Chirp fled from the cage, White Thrush pursued, teeth bared. The males also pursued, clucking and purring in anticipation. The males won, and White Thrush sulked the rest of the day. Guinea pigs apparently have a well-developed heterosexual affectional system.

Sex behavior in the guinea pig has been intensively investigated, and there are exhaustive studies on what has been called the sex drive, but I know of no previous mention of or allusion to the guinea pig's heterosexual affectional system. No doubt this stems from the paradigm which has been established for research in this area.

In a typical experiment a male guinea pig and a female guinea pig in estrus are taken from their individual cages, dropped into a barren chamber, and observed for 15 minutes. In such a situation there is a high probability that something is going to happen and that it will happen rapidly and repeatedly. The thing that happens will be reliable and valid, and all that one needs to do to score it is to count. It is my suggestion that from this time onward it be known as the "flesh count." Sometimes I wonder how men and women would behave if they were dropped naked into a barren chamber with full realization that they had only fifteen minutes to take advantage of the opportunities offered them. No doubt there would be individual differences, but we would obtain little information on the human heterosexual affectional system from such an experiment.

Sex is not an adventitious act. It is not here today and gone tomorrow. It starts with the cradle, and as a part of the human tragedy it wanes before the grave. We have traced and are tracing the development of the heterosexual affectional system in monkeys.

We believe that the heterosexual affectional system in the rhesus monkey, like all the other affectional systems, goes through a series of developmental stages—an infantile heterosexual stage, a preadolescent stage, and an adolescent and mature heterosexual stage. Although these stages are in considerable part overlapping and cannot be sharply differentiated in time, we would think of the infantile stage as lasting throughout the first year and being characterized by inadequate and often inappropriate sexual play and posturing. The preadolescent stage, beginning in the second year and ending in the third year in the female and the fourth year in the male, is characterized by adequate and appropriate sexual play and posturing, but incompleteness. The adolescent and adult stage is characterized by behaviors which are similar in form but give rise to productive outcomes which are also reproductive.

Since in this paper sex is an unavoidable issue, we present illustrations of normal adult macaque monkey sex behavior. Sexual invitation may be initiated by the female, as in Figure 1, by a present pattern with buttocks oriented toward the male, tail elevated, and the female looking backward with a fear-grimace (not threat) pattern involving flattened ears and lip smacking. As you can see, this pattern need not involve rape nor even rush on the part of the male. The male may also solicit, as in the case of the animal in the foreground of Figure 2; this animal has assumed a posture soliciting either grooming or more intimate favors. These patterns seldom elicit violent, uncontrolled, reflex behaviors. Normal male and female overt sex behavior is shown in Figure 3, the male having assumed the complex sex posture involving ankle clasp, dorsoventral mounting, and clasp of the female's buttocks. The partner demonstrates the complete female sexual pattern of elevating the buttocks, lowering the head, and looking backward. There have been millions of rhesus monkeys for millions of years, and there will be more in the future.

We have traced the development of the infantile heterosexual stage during the first year of life in two test situations using observational techniques. One is our playroom, illustrated in Figure 4, which consists of a room 8 ft. high with 36 feet of floor space. In this room are a platform, ladder, revolving wheel, and flying rings to encourage the infants' adaptation to a three-dimensional world, and there is an assortment of puzzles and toys for quieter activities. Two groups of four infants each, half of each group male and half female, have been observed in the playroom daily over many months. The second apparatus is shown in Figure 5. This is the playpen situation, and it consists of four large living cages and adjoining pens. Each living cage houses a mother and infant, and a three-inch by five-inch opening in the wall between cage and playpen units enables the infants to leave the home cage at any time but restrains the mothers. The playpen units are separated by wire-mesh panels which are removed one or two hours a day to allow the infants to interact in pairs during the first 180 days and both in pairs and in groups of four during the next half-year of life. Again, we are referring to data gathered from two playpen setups, each housing four infants and their real or surrogate mothers. Insofar as the infantile heterosexual stage is concerned, it makes little or no difference from which situation we take our data.

The outstanding finding in both the playroom and playpen is that male and female infants show differences in sex behavior from the second month of life onward. The males show earlier and more frequent sex behavior than do females, and there are differences in the patterns displayed by the sexes. The males almost never assume the female sex-posture patterns, even in the earliest months. The females, on the other hand, sometimes display the male pattern of sex posturing, but this is infrequent after ten months of age. Predominantly, females show the female pattern and exceptional instances are to other females, not males. Frequency of sex behavior for both males and females increases progressively with age. There is no latency period—except when the monkeys are very tired.

The early infantile sexual behaviors are fragmentary, transient, and involve little more than passivity by the female and disoriented grasping and thrusting by the male. Thus, the male may thrust at the companion's head in a completely disoriented manner or laterally across the midline of the body, as in Figure 6. However, it is our opinion that these behaviors are more polymorphous than perverse.

Thus, as soon as the sexual responses can be observed and measured, male and female sexual behaviors differ in form. Furthermore, there are many other behaviors which differ between males and females as soon as they can be observed and measured. Figure 7 shows the development of threat responses by males and females in the playroom, and these differences are not only statistically significant, but they also have face validity. Analysis of this behavior shows that males threaten other males and females but that females are innately blessed with better manners; in particular, little girl monkeys do not threaten little boy monkeys.

The withdrawal pattern—retreat when confronted by another monkey—is graphed for the playroom in Figure 8, and the significance is obvious. Females evince a much higher incidence of passive responses, which are characterized by immobility with buttocks oriented toward the male and head averted, and a similar pattern, rigidity, in which the body is stiffened and fixed.

In all probability the withdrawal and passivity behavior of the female and the forceful behavior of the male gradually lead to the development of normal sex behaviors. The tendency for the female to orient away from the male and for the male to clasp and tussle at the female's buttocks predisposes the consorts to assume the proper positions. The development of the dorsally oriented male sex-behavior pattern as observed in the playroom situation is shown in Figure 9 and may be described as a composite yearning and learning curve.

Infant male and female monkeys show clear-cut differences in behavior of far greater social significance than neonatal and infantile sex responses. Grooming patterns, which are basic to macaque socialization, show late maturation, but as is seen in Figure 10, when they appear, they sharply differentiate the two sexes. Caressing is both a property and prerogative of the females. Basic to normal macaque socialization is the infant-infant or peer-peer affectional system, and this arises out of and is dependent upon the play patterns which we have described elsewhere and only mention here. As is shown in the solid lines of Figure 11, play behavior in

Fig. 4: *Playroom test situation.*

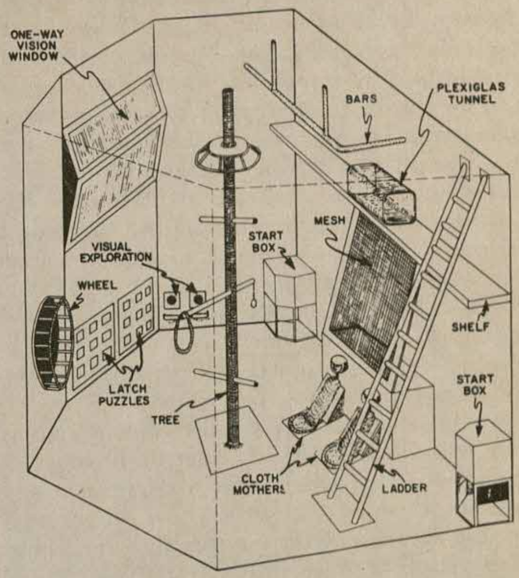
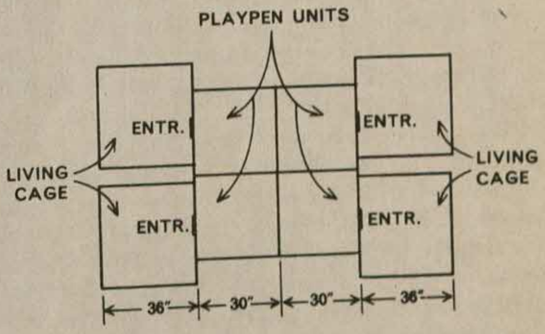


Fig. 5: *Playpen test situation.*



the playroom is typically initiated by males, seldom by females. However, let us not belittle the female, for they also serve who only stand and wait. Contact play is far more frequent among the males than the females and is almost invariably initiated by the males. Playpen data graphed in Figure 12 show that real rough-and-tumble play is strictly for the boys.

I am convinced that these data have almost total generality to man. Several months ago I was present at a school picnic attended by 25 second-graders and their parents. While the parents sat and the girls stood around or skipped about hand in hand, 13 boys tackled and wrestled, chased and retreated. No little girl chased any little boy, but some little boys chased some little girls. Human beings have been here for two million years, and they'll probably be here two million more.

These secondary sex-behavior differences probably exist throughout the primate order, and, moreover, they are innately determined biological differences regardless of any cultural overlap. Because of their nature they tend automatically to produce sexual segregation during middle and later childhood, but fortunately this separation is neither complete nor permanent. Behavioral differences may very well make it easy through cultural means to impose a sexual latency period in the human being from childhood to puberty. We emphasize the fact that the latency period is not a biological stage in which primary sex behavior is suppressed, but a cultural stage built upon secondary behavioral differences.

We believe that our data offer convincing evidence that sex behaviors differ in large part because of genetic factors. However, we claim no originality for the discovery of intersex behavioral differences. In 1759 Laurence Sterne in his book *Tristram Shandy* described male and female differences at the most critical period in Tristram Shandy's development; indeed, it would not be possible to conceive of a more critical period.

"*Pray, my dear,* quoth my mother, *have you not forgot to wind up the clock?*—— *Good G——!* cried my father, making an exclamation, but taking care to moderate his voice at the same time—— *Did ever woman, since the creation of the world, interrupt a man with such a silly question?*"[1]

Men and women have differed in the past and they will differ in the future.

It is possible that the listener has been dismayed by the frequent reference to sex and the relatively infrequent reference to affection. Out of these infantile behavior patterns, both sexual and nonsexual, develop the affectional bonds and the social ordering that appear to be important or even essential to the full development of the heterosexual affectional system of macaques. Traumatic affectional errors, both transient and prolonged, may have devastating effects upon subsequent social and sexual behaviors.

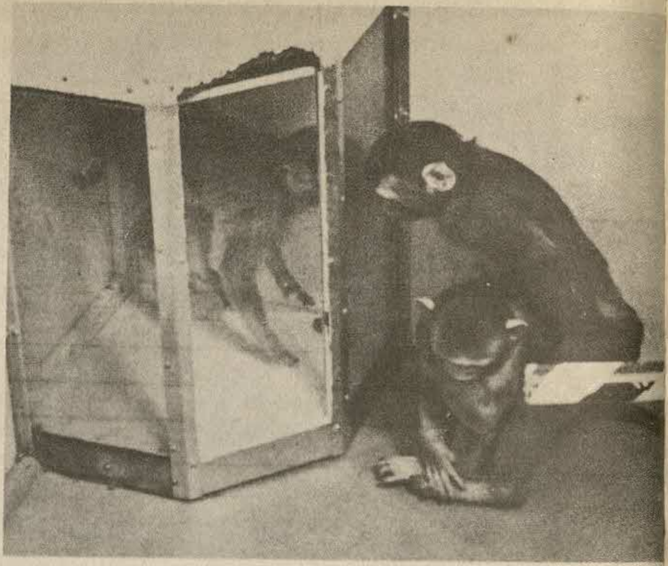


Fig. 6: *Immature male and female sexual posturing, playroom observation.*

For some years we have been attempting to establish experimental neuroses in infant monkeys by having them live on unfriendly and inconsistent mother surrogates. One preparation was a rejecting mother that on schedule or demand separated her baby when a wire frame embedded in her spun-nylon covering was displaced violently upward and backward. The baby was disturbed, but as soon as the frame was returned to its resting position, the baby returned to cling to its surrogate mother as tightly as ever. Next we developed an air-blast mother with a series of nozzles down the entire center of her body which released compressed air under high pressure—an extremely noxious stimulus to monkeys. The blasted baby never even left the mother, but in its moments of agony and duress, clung more and more tightly to the unworthy mother. Where else can a baby get protection? Apparently our infant had never read Neal Miller's theory that avoidance gradients are precipitous and approach gradients gradual and tenuous, for love conquered all.

---

[1] Sterne, Laurence. *The Life and Opinions of Tristram Shandy, Gentleman.* J. A. Work (Ed.). New York: The Odyssey Press, 1940. p. 6.

We next devised a shaking mother, which on schedule or demand shook her infant with unconscionable violence until its teeth chattered. The infant endured its tribulations by clinging more and more tightly. At the present time we believe we may be on the threshold of success through Jay Mowbray's creation of the porcupine mother, which extrudes brass spikes all over its ventral surface. Preliminary studies on two infants suggest that they are emotionally disturbed. Whether or not we eventually succeed, the fact remains that babies are reluctant to develop experimental neuroses, and at one time we even wondered if this were possible.

During the time that we were producing these evil mothers, we observed the monkeys which we had separated from their mothers at birth and raised under various mothered and non-mothered conditions. The first 47 baby monkeys were raised during the first year of life in wire cages so arranged that the infants could see and hear and call to other infants but not contact them. Now they are five to seven years old and sexually mature. As month after month and year after year have passed, these monkeys have appeared to be less and less normal. We have seen them sitting in their cages strangely mute, staring fixedly into space, relatively indifferent to people and other monkeys. Some clutch their heads in both hands and rock back and forth—the autistic behavior pattern that we have seen in babies raised on wire surrogates. Others, when approached or even left alone, go into violent frenzies of rage, grasping and tearing at their legs with such fury that they sometimes require medical care.

Eventually we realized that we had a laboratory full of neurotic monkeys. We had failed to produce neurotic monkeys by thoughtful planning and creative research, but we had succeeded in producing neurotic monkeys through misadventure. To err is human.

Because of housing pressures some of these monkeys and many of our surrogate-raised monkeys lived in pairs for several years while growing to sexual maturity, but we have seldom seen normal sex behavior, and we certainly have not had the validating criterion of newborn baby monkeys. Instead, these monkeys treat each other like brother and sister, proving that two can live in complete propinquity with perfect propriety as long as no one cares.

Fig. 7: *Frequency of threat responses by males and females in the playroom.*

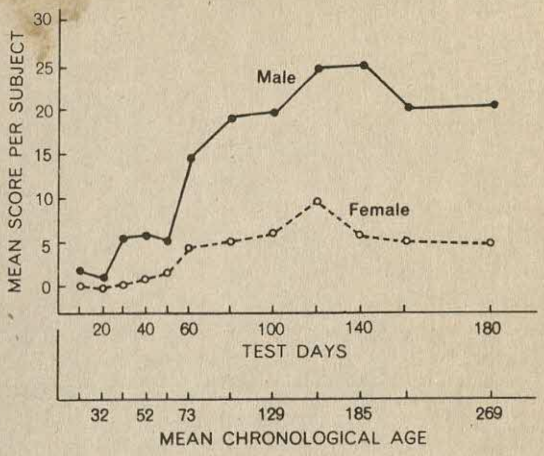
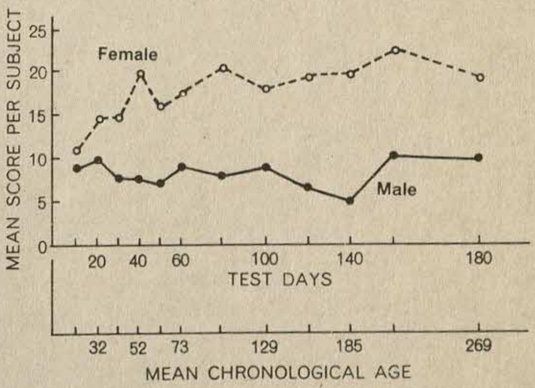


Fig. 8: *Frequency of withdrawal responses by males and females in the playroom.*



Their reason for being, as we saw it, was to produce babies for our researches, and so at this point we deliberately initiated a breeding program which was frighteningly unsuccessful. When the older, wire-cage-raised males were paired with the females at the peak of estrus, the introduction led only to fighting, so violent and vicious that separation was essential to survival. In no case was there any indication of normal sex behavior. Frequently the females were the aggressors; even the normal praying mantis waits until the sex act is completed.

Pairing such cloth-surrogate-raised monkeys

Fig. 9: *Percentage of all male mounts (immature and mature) in the playroom that shows dorsal orientation (mature pattern).*

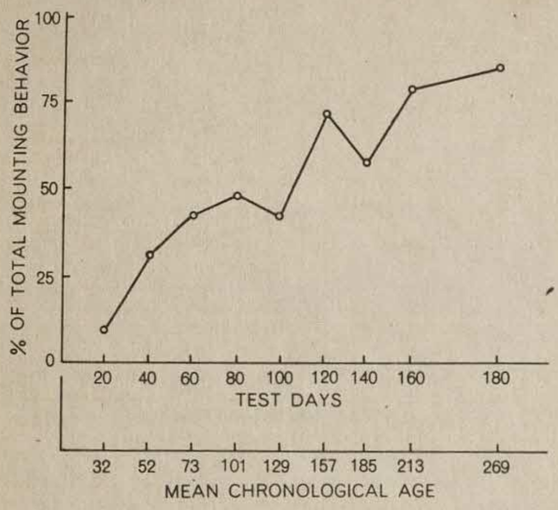
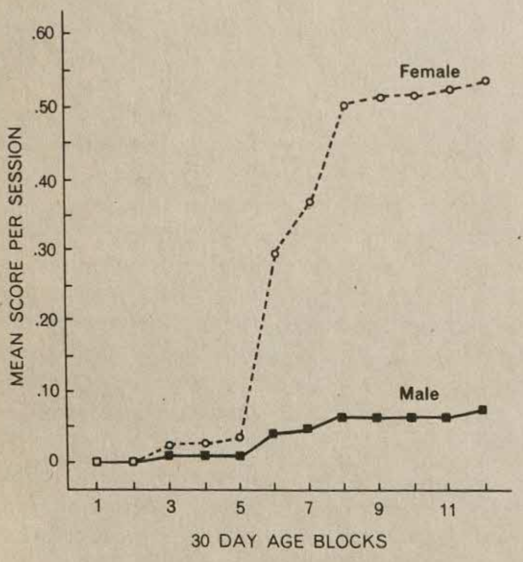


Fig. 10: *Frequency of grooming responses made by males and females in the playroom.*



as were sexually mature gave little better end results. Violent aggression was not the rule, and there was attempted sex behavior, but it was unreproductive since both the male and female behaviors were of the infantile type we have already described.

At this point we took the 17 oldest of our cage-raised animals, females showing consistent estrous cycles and males obviously mature, and engaged in an intensive re-education program, pairing the females with our most experienced, patient, and gentle males, and the males with our most eager, amiable, and successful breeding females. When the laboratory-bred females were smaller than the sophisticated males, the girls would back away and sit down facing the males, looking appealingly at these would-be consorts. Their hearts were in the right place, but nothing else was. When the females were larger than the males, we can only hope that they misunderstood the males' intentions, for after a brief period of courtship, they would attack and maul the ill-fated male. Females show no respect for a male they can dominate.

The training program for the males was equally unsatisfactory. They approached the females with a blind enthusiasm, but it was a misdirected enthusiasm. Frequently the males would grasp the females by the side of the body and thrust laterally, leaving them working at cross purposes with reality. Even the most persistent attempts by these females to set the boys straight came to naught. Finally, these females either stared at the males with complete contempt or attacked them in utter frustration. It became obvious that they, like their human counterpart, prefer maturer men. We realized then that we had established, not a program of breeding, but a program of brooding.

We had in fact been warned. Our first seven laboratory-born babies were raised in individual cages while being trained on a learning test battery. William Mason planned to test their social behaviors subsequently, and great care had been taken to keep the babies socially isolated and to prevent any physical contacts. Neonatal baby monkeys require 24-hour-a-day care, and infant monkeys need ministrations beyond a 40-hour week. We had assigned the evening care to Kathy, a maternal bit of fluff who had worked for several years as a monkey tester while studying to become an elementary school teacher.

Checking on his wards one night near 10 P.M., Mason found Kathy sitting on the floor surrounded by seven baby monkeys, all eight of the primates playing happily together. Before the horrified scientist could express his outrage, Kathy had risen to her full height of five feet two. Already anticipating the carping criticisms which he was formulating, she shook her finger in his face and spoke with conviction: "Dr. Mason, I'm an education student and I know that it is improper and immoral to blight the social development of little children. I am right and you are wrong!"

Although we were angry with Kathy, we did think there was a certain humor in the situation and we did not worry about our monkeys. We simply transferred Kathy to an office job. Alas, she could not have been more right and we could not have been more wrong! We have already described the social-sexual life of these 7 monkeys and the next 40 to come.

Two years later we had more than theoretical reasons to be disturbed because Mason tested a group of these isolation-raised monkeys, then between 2.5 and 3.5 years of age, and found evidence of severe social abnormalities, which might be described as a sociopathic syndrome. He matched the laboratory-raised monkeys on the basis of weight and dentition patterns with monkeys that had been born and raised in the wild for the first 12 to 18 months, then captured and subjected to various kinds of housing and caging treatments for the next year or two. In the test situations the laboratory-raised monkeys, as compared with feral monkeys, showed infantile sexual behavior, absence of grooming, exaggerated aggression, and absence of affectional interaction as measured by cooperation.

We are now quite certain that this sociopathic syndrome does not stem from the fact that the baby monkeys were raised in the laboratory but from *how* they were raised in the laboratory. Our infants raised in the laboratory by real monkey mothers and permitted opportunity for the development of normal infant-infant affection demonstrate normal male and female sexual behavior when they enter the second year of life. Furthermore, our playroom and playpen studies show that infant monkeys raised on cloth mothers but given the opportunity to form normal infant-infant affectional patterns, also develop normal sexual responses.

In a desperate attempt to assist a group of 18 three- to four-year-old cloth-surrogate-raised monkeys, half of them males and half females, we engaged in a group-psychotherapy program, placing these animals for two months on the monkey island in the Madison Zoo. Their summer vacation on the enchanted island was not without avail, and social grooming responses rapidly developed and were frequent in occurrence. After a few days of misunderstanding, patterns of social ordering developed, and a number of males and females developed friendship patterns. Unfortunately, sexual behavior was infrequent, and the behavior that was observed was completely inadequate—at least from our point of view. In desperation we finally introduced our most experienced, most patient, and most kindly breeding male, Smiley (the male in Figures 1 and 2), and he rapidly established himself as king of the island and prepared to take full advantage of the wealth

Fig. 11: *Frequency of play-initiations by males and females to monkeys of the same (male-male, female-female) and other sex (male-female, female-male). Observations are from the playroom.*

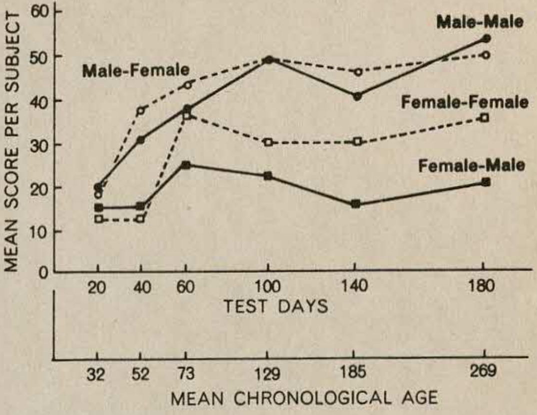
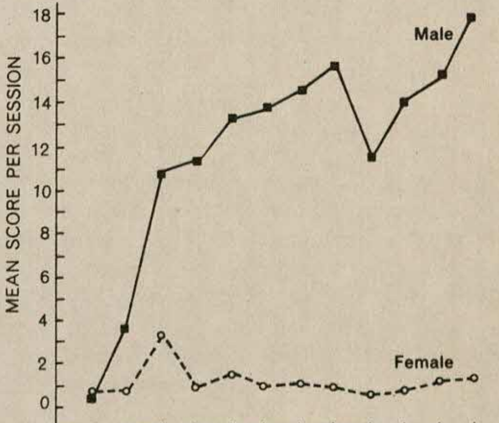


Fig. 12: *Frequency of occurrence of "rough-and-tumble" play for two males and two females in the playroom through the first year of life.*



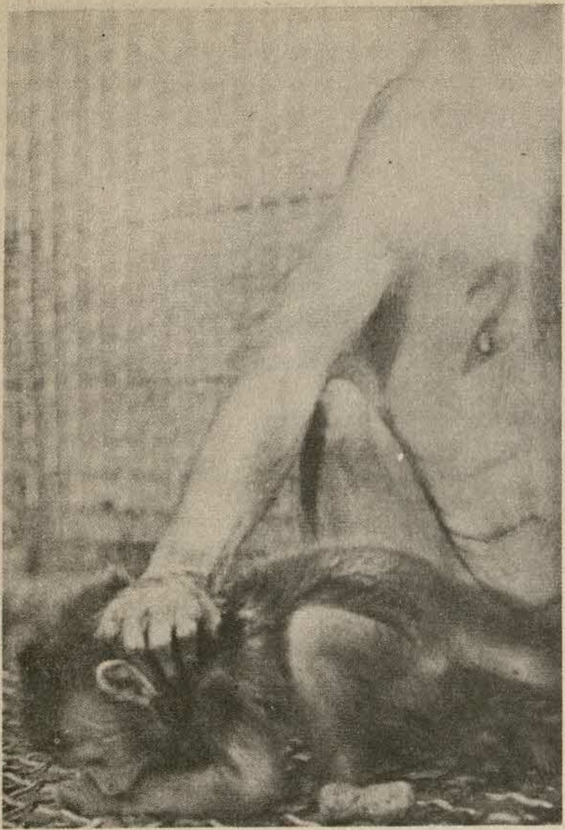


Fig. 13: *Typical behavior of unmothered mother toward her infant. Mother is looking upward while crushing her baby against the cage floor.*

of opportunity which surrounded him. Fortunately, the traumatic experiences he encountered with unreceptive females have left no apparent permanent emotional scars, and now that he has been returned to our laboratory breeding colony, he is again making an important contribution to our research program. If normal sexual behavior occurred, no member of our observational team ever saw it, and had a female become pregnant, we would have believed in parthenogenesis.

But let us return to the monkeys that we left on the island and the older ones that we left in their cages. A year has passed, and the frustrations that both we and our monkeys experienced are in some small part nothing but a memory. We constructed larger and more comfortable breeding cages, and we designed a very large experimental breeding room 8 feet by 8 feet by 8 feet in size with appropriate platforms and a six-foot tree. Apparently we designed successful seraglios for I can report that not all love's labors have been lost. It does appear that the males are completely expendable unless they can be used in a program of artificial insemination. Certainly we can find no evidence that there is a destiny that shapes their ends unless some Skinnerite can help us with the shaping process. We have, however, had better success with some of the females, particularly the females raised on cloth surrogates.

Even so, one of the wire-cage-raised females is a mother and another is pregnant. Three cloth-surrogate females are mothers and four or five are expectant. We give all the credit to three breeding males. One, Smiley, does not take "no" for an answer. Smiley has a way with females. Patient, gentle, and persuasive, he has overcome more than one planned program of passive resistance. One female did not become pregnant until the fifth successive month of training. Month after month she has changed, and now she is mad about the boy. Male No. 342 behaves very much like Smiley. Even when females threaten him, he does not harm them. Given time, he has been able to overcome more than one reluctant dragon, and he is master of the power of positive suggestion.

Breeding male No. 496 has helped us greatly, particularly with the younger, cloth-surrogate-raised females. His approach differs from that of Smiley and No. 342. His technique transcends seduction, and in contract bridge terms it may be described as an approach-forcing system.

Combining our human and male-monkey talents, we are winning the good fight and imparting to naive and even resistant female monkeys the priceless gift of motherhood. Possibly it is a Pyrrhic victory. As every scientist knows, the solution of one scientific problem inevitably leads to another, and this is our fate (Figure 13). Month after month female monkeys that never knew a real mother, themselves become mothers —helpless, hopeless, heartless mothers devoid, or almost devoid, of any maternal feeling.

PART III

EARLY CHILDHOOD

THE PARENT-CHILD RELATIONSHIP

COGNITIVE DEVELOPMENT

PERSONALITY DEVELOPMENT

S.C.E.R.T., West Bengal
Library
Calcutta
25/3, B. C. Road

THE PARENT-CHILD RELATIONSHIP

20

The Influence of Degree of Flexibility in Maternal Child Care Practices on Early Child Behavior

Ethelyn H. Klatskin, Edith B. Jackson, and Louise C. Wilkin

In this paper on some of the results of their longitudinal study of the Yale Rooming-In Project, Klatskin and her colleagues report on the relative effects of different child-rearing practices on the specific behaviors of the child during the first three years. Again the importance of maternal attitudes is stressed. The reader should note that both longitudinal material and a cross-sectional study are combined in one article.

The present paper is a report on the analysis of a selected group of cases from the files of the Yale Rooming-In Research Project, a longitudinal study of parent-child relationship focused on observing the effect of certain specific parental child care practices on child behavior. The research program has utilized a team approach to collect data on the developing family unit during the prenatal period, in the hospital during the puerperium, and during the first three years of the child's life. The details of the development of the Project and the methodology of data collection and organization have been reported elsewhere (Klatskin & Jackson, 1955).

The study had its inception during a period which marked a cultural shift from rigid to permissive practices in child training. For the previous twenty years, medical trends and Watsonian behaviorism had influenced pediatricians and educators to advocate inflexible techniques in child training. Beginning around 1935, increasing familiarity with Freudian analytic tenets and the necessity to deal with an apparently increasing number of behavior problems led a number of child psychiatrists to propose the use of more flexible practices in early child care. This orientation was in large part due to the conviction that many of the then observed problems in early child development were the result of rigid parental child care practices. The ensuing twenty years have witnessed an almost complete shift in popularity from inflexible to flexible practices, both as advocated by specialists in the field and as accepted by the general public.

In neither the preceding period of rigidity nor the current period of permissiveness, however, has any body of statistical evidence been presented to substantiate the claims of the proponents of either type of child training. The present research was organized to amass just such a body of data. In general, the Project has been predicated on the hypothesis that a demonstrable relationship can be shown to exist between parental child care practices on the one hand

---

From *American Journal of Orthopsychiatry*, 1956, *26*: 79–93.
Copyright, the American Orthopsychiatric Association, Inc. Reproduced by permission.

Table 1: *Parents' education**

		Mother's Education			
		Grammar School	*High School*	*College*	*Advanced Degree*
Father's Education	Grammar School	3	1		
	High School		16	6	
	College			6	
	Advanced Degree			13	5

** Classified according to attendance at or completion of the given level.*

and the child's behavior on the other. Specifically, the hypothesis investigated was that either extreme of rigidity or permissiveness in parental practices would be associated with problem behavior in the child, within the area of the parental deviation.

SUBJECTS AND PROCEDURE

The subjects of the present study were 50 mothers and infants chosen from the families followed for three years through the research project. Two factors determined selection of this group: the mother's parity and the completeness of the case history. In order not to obscure the initial parent-child interaction, the records of primiparae were chosen for analysis. The first 50 records were then selected which met the requirements of including all of the major research procedures: office visit and psychological examination, social worker's home visit, and parental report by questionnaire, at one, two and three years.

The socioeconomic composition of this subgroup is given in Tables 1 and 2. In defining the occupational groupings, the categories of the United States Census were used. In both tables, students in training were classified according to educational and occupational expectation. In Table 2, the mother was classified according to either previous or present type of employment; actually, only 11 of the 50 mothers studied were working during any period of the three year follow-up.

The tables show that approximately half of the group was composed of families in which both parents had college or advanced degrees, and that in one-third of the group both parents were at some time in professional or managerial positions. There was no representation from the highest or lowest socioeconomic groups. The relationships within the sample can best be described in terms of Hollingshead's socioeconomic grouping for the New Haven area (1953): Class II (professional and managerial), 19; Class III (white collar and skilled workers), 19; and Class IV (semiskilled and unskilled workers), 12.

A rating scale technique of analysis was used to reduce the data contained in the fifty case histories to statistically manipulable form.[1] Independent scales were constructed for rating both parental and child behavior in the areas of feeding, sleeping, toileting, and socialization. Additional scales were constructed for the data on mother's adjustment to the maternal role, and for evidences of emotional disturbance in the child, other than those which might be included in the four behavior areas. Since the data broke logically at the child's first, second, and third birthdays because of the timing of the research procedures, each year was scaled separately.

In scaling the areas of feeding, sleeping, toileting, and socialization, the major variable for the mother's behavior was defined as degree of rigidity in handling the child, and for the child's behavior, degree of problem. The scales of maternal behavior ranged from extreme rigidity (point 1) at one end of the scale to extreme permissiveness at the other (point 5), with flexible handling as the mid-point (point 3). The scales of child behavior ranged from severe problem at one end of the scale (point 1) to absence of problem at the other (point 5), with "problem within normal limits" as the mid-point (point 3). Either extreme of the scale of maternal behavior was assumed to represent deviant handling leading to the development of problems in the child, while the mid-point of the maternal behavior scale was assumed to represent optimal handling associated with normal development.

---

[1] For details of method of construction of these scales, see Klatskin and Jackson (1955).

Table 2: *Parents' occupation*

		Mother's Occupation				
		Professional-Managerial	*White Collar*	*Skilled*	*Semi-skilled*	*None*
Father's Occupation†	Professional-Managerial	10	9			
	White Collar	2	9	1	1	2
	Skilled	1	2			1
	Semiskilled	1	2		1	2
	Unskilled		2		2	2

* *Classified according to present or previous occupation.*
† *Classified according to present occupation or occupational expectation.*

The one exception to this assumption occurred in the evaluation of maternal behavior in toileting during the first year. In this instance, any rating below the mid-point of the scale was regarded as rigid and any rating above the mid-point as optimal.

The scale for rating the mother's adjustment to the maternal role ranged from complete rejection at one end, to complete absorption at the other, with "balanced adjustment" as the mid-point. The scale for other evidences of emotional maladjustment in the child ranged from severe and/or numerous problems at one end to absence of problems at the other, with mild and/or infrequent problems as the mid-point.

Two considerations determined the application of the scales to the evaluation of the case records: the necessity to have an independent rating for mother and child; and to have a rating which would be both reliable and valid. To meet these requirements, six raters were used, who were assigned in groups of three to either mother or child. Both assignment to a group and to a subject was successively randomized among the six raters. In this way, three independent judgments for each segment of behavior were available for both mother and child, except in those instances where insufficient information was available in the record. A reliable judgment was assumed to be one on which two or more raters agreed, and only these were recorded for subsequent correlation. The assignment of raters to successive groupings provided a measure of the extent to which any given rater consistently deviated from the other two, it being assumed that the ratings were more valid if no significant variability occurred within the group. That is, if no rater was found to deviate consistently from the others, the probability would be increased that all were using the same bases for their judgments. The alternation between rating mother and child was planned to focus the raters on the scales themselves as a reference point, rather than on the consecutive records. No significant differences were found between the raters in either number of times they deviated from the group, or number of times they reported insufficient data.

RESULTS

The initial step in analysis of the data was to plot the distribution of the consensus judgments for both mother and child for each segment of the four behavior areas. With two exceptions, both in the area of toileting, the ratings tended to assume a normal distribution around the mid-point (3) of the scale. In the area of toileting, when the mothers' ratings were plotted, it was found that at the first year, only 15 mothers fell below the mid-point of the scale, and at the third year, no information was returned by 21 of the group. In these 21 cases, the mothers replied that the child was trained, although according to the scale for the child's behavior, only 11 were rated as being trained. The handling of these data therefore constituted an exception to the type of analysis described below.

Since the various points on the mother's and child's scales did not correspond directly to each other (as the mother's scale represented a progression from undesirable rigidity through optimal handling to undesirable overpermissiveness, while the child's scale represented a progression from undesirable to desirable behavior), this fact had to be taken into account in analyzing the relationship between the two. A chi-square rather than correlational analysis was therefore planned. The mothers' scales represented a fairly absolute judgment of deviant behavior, since the extremes of overpermissiveness and rigidity for the various areas were defined so as to take into account marked departures from the normal attitude. In each area, approximately half of the group (between 43% and 60%) fell into the optimal category, while the position of the remaining half varied somewhat according to the behavior area rated. The mothers were therefore initially divided into those obtaining ratings of 1 or 2 (rigid), ratings of 3 (optimal), and ratings of 4 or 5 (overpermissive). The child's scale, however, was found to indicate a more relative degree of disturbance. When the ratings for the various areas were plotted, all tended to assume a normal distribution, but the mode varied slightly for the different years and variables. Thus, the modal score for socialization, first year, was 4.0; for socialization, second year, it fell to 3.5; and for socialization, third year, it was 3.0. In order to maintain a consistent reference point, ratings for the child in each area were therefore dichotomized at the median of the group, dividing the children into those with more or less problem behavior in a given area. Dividing the children's group in this way arbitrarily resulted in a range of incidence of "problem" behavior for the various areas of 33 to 53 per cent. It should be kept in mind throughout the subsequent presentation of results that when "presence of problem behavior" or "absence of problem behavior" is discussed, the child is being assessed relative to his position in the present group of subjects, and not in terms of more absolute standards which might be applied to "problem children" in a child guidance clinic.

The data for maternal attitude toward toileting in the first and third years were treated as follows: at the first year, the mothers were dichotomized into those falling below the mid-point of the scale and those at or above the mid-point. For the third year, in view of the lack of data, the mothers' ratings at two years were correlated with the child's behavior at three years.

Chi-square was then computed for all the relationships, with the exception of first year toileting, on the basis of a six-cell table, trichotomizing the mothers' group into rigid, optimal, and overpermissive, and dichotomizing the group of children into those above and below the median incidence of problem behavior. Of the 12 possible relationships, 7 were found to be significant at or below the .01 level. However, it was apparent that because of the small number of cases, some relationships between deviant maternal behavior and the development of problem behavior in the child were obscured. The chi-squares were therefore recalculated on the basis of a four-cell table, dichotomizing the mother's handling into optimal and deviant, without regard to direction. The results of these latter calculations are given in Table 3; those values of chi-square significant when computed on the basis of the six-cell tables are also indicated. Because of the number of significance tests calculated in analyzing the data, the criterion for a "significant" relationship was set as probability of occurrence by chance less than twice in a hundred. Consequently, only those relationships significant at or below the .02 level of confidence will be reported here.

The direction of the maternal deviation associated with "problem" behavior in the child is indicated in Table 4. It is apparent from this table that a consistent relationship exists be-

Table 3: *Significant values of chi-square between deviant or optimal maternal handling and presence or absence of problem behavior in the child*

Year	Area			
	Feeding	Sleeping	Toileting	Socialization
1	—	χ^2—11.77* $p < .001$	χ^2— 9.91 $p < .01$	—
2	χ^2—13.00 $p < .001$	χ^2— 8.42 $p < .01$	χ^2—12.74* $p < .001$	χ^2— 7.18* $p < .01$
3	χ^2— 5.75 $p < .02$	χ^2—15.25* $p < .001$	χ^2—13.97* $p < .001$	χ^2—12.58* $p < .001$

* $p < .01$ for six-cell table.

Table 4: *Direction of maternal deviation significantly associated with problem behavior in the child*

Year	Area			
	Feeding	Sleeping	Toileting	Socialization
1		Overpermissive		
2	Rigid	Overpermissive	Both	Rigid
3	Rigid	Overpermissive	Both	Rigid

tween type of maternal deviation and the behavior area in which this is shown.

Tables 3 and 4 indicate that the relationship between maternal behavior and child behavior tends to be a discrete one, in that the child develops "problems" in response to specific deviations from optimal handling. As a further check on this conclusion, interarea relationships were tested to determine whether maternal deviation in one area was associated with problem behavior in the child in a different area. For example, maternal behavior in feeding was related to the child's behavior in sleeping. No significant relationships of this type were found.

The question might still be raised as to whether some over-all factor of adjustment or maladjustment in the child resulted in the presence or absence of problem behavior. The child's ratings on the scale of "Other Evidences of Emotional Maladjustment," which had been designed to assess the child's general emotional stability, were therefore analyzed in relation to his ratings in each of the four areas for the three years. The only relationships found ($p < .02$) were in the area of socialization, where during the first and third years, children with "problems" in socialization tended to fall below the median of the group in evidencing "emotional maladjustment." Because of the possibility that this relationship might have arisen from use of the same data in rating both areas, the ratings at the third year were checked for such duplication. Of the 307 individual items of behavior listed by the judges as the basis for their ratings of the child's social adjustment in the 50 cases, only 39, or 13 per cent, were repeated as evidence for additional emotional maladjustment.

Finally, major environmental events in the child's life were related to his behavior in the four areas as well as to his rating with respect to the "emotional maladjustment" scale. The factors studied were: intelligence, dividing the children into below average (5 cases), average (30 cases), and above average (15 cases); duration of breast feeding, dividing the group at the median duration (3 months); birth of a sibling (20 cases); nursery school attendance in the third year (16 cases); and severe and/or recurrent illness (10 cases). The only positive relationship found was that between birth of a sibling during the second year and the presence of greater emotional maladjustment ($x^2 - 8.75$, $p < .01$).

Since it appeared from the foregoing data that the mother's handling of the child was the major variable in influencing his behavior, the next step in data analysis was to investigate what factors might be influencing the maternal attitude. These fell into three general categories—the mother's general emotional stability, as measured by her adjustment to the maternal role; environmental factors current in the home situation; and the mother's social class membership. When the mother's adjustment to the maternal role was related to her behavior toward the child, significant relationships were found in the areas of handling of sleep and socialization. The chi-square values for these, based on dichotomizing the mother's adjustment as deviant or balanced, and her handling of the child as deviant or optimal, are given in Table 5. In the case of sleep, rejection of the maternal role tended to be associated with rigid practices in this area, and absorption tended to be associated with overpermissive practices. The most striking relationships, however, were found in the area of socialization. During the first year, of the 16 mothers judged as rejecting of the maternal role, 13 used rigid methods of socialization; in the second year, of 20 judged as rejecting, 17 used rigid practices; and in the third year, of 17 judged as rejecting, 15 used rigid practices.

An attempt was next made to evaluate the importance of environmental factors in influencing both actual handling of the child and the mother's adjustment to the maternal role. The following "deviant home situations" were selected for analysis: marked discrepancy in parental background, inadequate housing, prolonged illness of either parent, atypical employment

Table 5: *Relationship between attitude toward maternal role and maternal behavior*

Year	Area			
	Feeding	*Sleeping*	*Toileting*	*Socialization*
1	—	χ^2— 8.16 $p < .01$	—	χ^2—24.42 $p < .001$
2	—	χ^2—11.94 $p < .01$	—	χ^2—12.09 $p < .001$
3	—	χ^2— 5.92 $p < .02$	—	χ^2—14.28 $p < .001$

hours for father, and mother working or out of the home. It was found, however, that these variables occurred too infrequently to permit statistical analysis, and though the presence of any one or constellation of them was of obvious importance for the dynamics of the individual case, such analysis is not within the scope of the present paper.

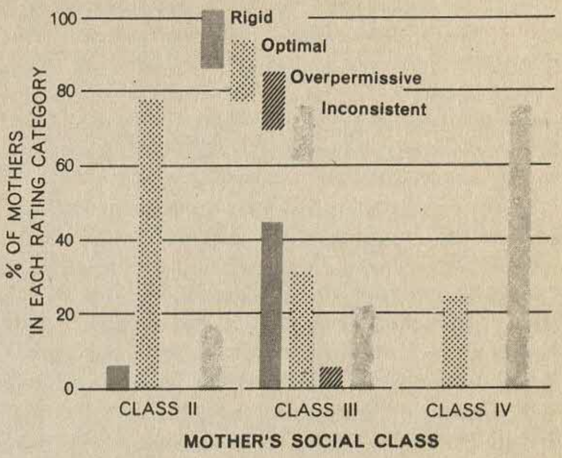
Because much emphasis has been placed on the mother's social class membership as a determinant of maternal child care practices, the mothers' behavior in the four areas was analyzed for the three classes represented in the present group of subjects. Here, definite trends were observed which, because of the small number of subjects, were not statistically significant. A summary of these data is presented in Figure 1, which was arrived at in the following manner. Since 12 attitude ratings were available for each mother (one for each of the four areas for each ofthe three years), a total score could be assigned to her based on the category into which the majority of the ratings fell. For purposes of the present summary, a mother was assigned that score into which two-thirds (8 out of the 12) of her ratings fell. On this basis, she could be classed as rigid, optimal, overpermissive, or inconsistent (no two-thirds majority). It will be seen from Figure 1 that the majority of the Class II mothers fell into the optimal category, the majority of the Class III mothers fell into the rigid category, and the majority of the Class IV mothers showed no consistent pattern, but varied in optimal, rigid, or overpermissive behavior from area to area.

Two questions which have been raised about the present data are 1) the relationship between the consistency of the mother's behavior and the development of problems in the child, and 2) the effect on the child of deviant maternal attitudes in those mothers who were judged as well adjusted to the maternal role as opposed to those mothers who were judged as rejecting or overabsorbed.

In order to assess the cumulative effect of maternal consistency or inconsistency, the data were analyzed for each of the four major behavior areas. The child's status at three years (above or below the median of his group with reference to degree of problem behavior) was related to the mother's behavior throughout the three years. The mothers were classed as consistently optimal, rigid, or overpermissive (those who received the same rating at all three years) and inconsistent (those who received a different rating for any or all of the three years). When the consistent mothers were contrasted with the inconsistent mothers, no difference was found in incidence of problem behavior in the children. However, when contrasts were drawn within the group of consistent mothers, significant differences were found in incidence of problem behavior between the children of consistently optimal mothers on the one hand and the children of consistently rigid or overpermissive mothers on the other. In each instance, the children of the consistently optimal mothers had a significantly lower incidence of problem behavior. The chi-square values for the four areas are as follows: feeding—4.24, $p < .05$; sleeping—8.69, $p < .01$; toileting—12.62, $p < .01$; and socialization—4.24, $p < .05$.

The hypothesis has been advanced that fewer problems would develop among children handled deviantly by mothers who were well adjusted to the maternal role than among those whose mothers were rejecting or overabsorbed. Because of the small number of cases in the present sample, and the high correlation obtained between maternal adjustment and optimal handling, it was not possible to contrast the effect on the child of deviant handling by the different types of mothers. It was hoped that this type of breakdown would yield information as to the effect on child development of overpermissiveness or rigidity in mothers who felt at ease in such practices (that is, those mothers adjusted to the maternal role), as opposed to deviation which was the outgrowth of superficial adherence to "popular" advice or represented rejection of the child (that is, those mothers overabsorbed or rejecting of the maternal role). Of the 50 cases, 14 mothers were judged as adjusted to the maternal role at all three years; 20 were judged as maladjusted at all three years; and in 16 cases the mother's type of adjustment varied from year to year. As might be anticipated from the data given in Table 5, the majority of the well-adjusted mothers showed optimal handling, and the majority of the maladjusted mothers showed deviant handling. Consequently, no contrasts were possible.

Fig. 1: *Relationship between mother's social class membership and her predominant type of child care practice.*



DISCUSSION

The present results show that child behavior during the first three years is more consistently influenced by maternal handling than by other environmental factors. However, in view of the stress which has been placed on the role played by early infancy experience in subsequent development, it is notable that fewer significant relationships were found within the first year than for the two subsequent years. Not only was no significant relationship found in the first year between maternal practice and child behavior in the areas of feeding and socialization, but when the mother's first year practices were related to the child's behavior in the second year to determine whether a delayed effect could be shown, no significant relationships appeared.

There is no ready explanation to account for the fact that in these areas the type of maternal handling during the first year does not here have the same obvious effect on the child's behavior as it does during the second and third. However, three possibilities suggest themselves. The first is that the present techniques were not sufficiently sensitive to detect greater individual variation among the children in these areas than was found during the first year. This is suggested by the fact that although in the first year about the same number of mothers fell into the categories of rigid, optimal, and overpermissive as during the second and third years, a higher proportion of the children fell above the mid-point of the scale (point 3) in the first year than during the second and third years. Using the present techniques of dividing the children's group at the median and thus making a relative judgment as to presence or absence of problem, in the area of feeding, 19 fell below the median and 27 above; in the area of socialization, 22 fell below the median and 26 above. However, dividing the group at the mid-point and thus making a more absolute judgment as to presence or absence of problem, in feeding only 5 fell below the mid-point, with 40 above it; in socialization, only 7 fell below the mid-point, with 42 above it. Thus, while it appears that this group of children were relatively problem free in these two areas during the first year, it is possible that if finer discriminations could have been made between them, a stronger relationship to type of maternal care might have been found. According to this hypothesis, incipient problem behavior actually existed among the children, but the research techniques were not sensitive enough to detect it.

An alternative explanation is that while this group of mothers showed variation in type of maternal handling, none showed the gross distortions in child care that have previously been correlated with the development of feeding and socialization problems during the first year. It is plausible to assume that, as in the area of toileting, overpermissiveness during the first year in feeding and socialization may not represent the gross mishandling that it does during the second and third years. On the other hand, it is possible that the judgment of rigidity made for the present mothers did not represent the same extreme of control that characterized child care practices 20 years ago. Both interpretations would decrease the probability of finding a significant relationship between maternal care and child behavior. According to this hypothesis, the group of children would be regarded as relatively problem free in these two areas because minor variations in handling among an essentially homogeneous group of mothers would be less likely to produce problem behavior during the first than during subsequent years.

A third possibility is that, because of the nature of the group and the amount of pediatric supervision received, less variation existed among the mothers' attitudes toward the child in the first year than in the second and third. It will be recalled that the present group of subjects were drawn from an "average" population of women routinely attending an obstetrical clinic. In all but 14 of the 50 cases, the baby was planned, and was presumably desired and responded to with affection and acceptance at his birth. During the first year, the mothers—either through attendance at a Child Health Conference of the Visiting Nurse Association or through a private pediatrician—received more intensive supervision than during the second and third years of the child's life. A situation therefore arose in which the mothers received more help in understanding the child when the needs to be met were the basic ones of feeding, sleeping, and protection from injury, than subsequently when he became more independent and active in exploring his environment and the cultural pressures for his "socialization" were increased. Although the mothers could be categorized as rigid, optimal, or overpermissive during the first year, it seems plausible to assume that less variation in underlying attitudes would be found during the child's infancy than later in his life, when his demands became more individual. As a result of the necessity for increased discipline, and for meeting the demands of society that he be toilet trained, learn appropriate techniques in eating, and acquire recognition of the distinction between his property and that of others, the degree of interaction between mother and child is heightened, and presumably a wider range of maternal attitudes are elicited. According to this hypothesis, the attitudes underlying the child care practices during the first year within this group would be more homogeneous than during the second and third years, when the mother's own anxieties and conflicts would be aroused by the child's increasing independence and by the increasing demands of society, and when, because of less intensive pediatric supervision, she would have to deal with him more in terms of her own individual personality and previous experiences.

Another striking aspect of the data which requires explanation is the discreteness of the relationships found between maternal handling and the child's reaction. It is apparent that poor handling in any given area is likely to be associated with "problem" behavior in the child in that area, but that this does not generalize in the child to other areas. It is also apparent that the type of maternal deviation varies in effect in the different areas of child behavior. Thus, in the area of sleep, overpermissiveness produces problems, while in the areas of feeding and socialization, rigidity produces problems. Here again, there are several possible explanations. One is that such specificity is due in part to the structure of the group. Thus, on the one hand, very few of the mothers were rated as falling at the extremes of the scales (points 1 and 5), so that the rigid and overpermissive groups were composed largely of mothers who received intermediate ratings between the midpoint and the extremes. If the mothers had shown more gross deviation from what would be regarded as acceptable child care practices, greater spread of effect might have been found among the children. This interpretation derives some support from the fact that few of the present group of children evidence any severe emotional disturbance of the kind that would lead to a pervasive disruption of behavior. The hypothesis that greater generalization of problem behavior would occur if the child were more generally disturbed is borne out by further analysis of the present data. If, at each year, the children are divided at the median of the scale of emotional maladjustment into those more or less emotionally maladjusted, the median number of problems is always greater in the group judged as more maladjusted. Thus, at the first and second years, the median number of problems for the more maladjusted group fell between 2 and 3, for the less maladjusted, between 1 and 2; at the third year, the median for the more maladjusted group again fell between 2 and 3, but for the less maladjusted, between 0 and 1.

It can also be hypothesized that the relationships between child behavior and parental child care practice may be more discrete at the time of origin of problems than later in the child's life. The observation that certain neurotic components of personality represent anxiety displaced from early childhood experience is undoubtedly a valid one, but it has been deduced for the most part from the treatment of older children and adults. The present direct observations of early childhood suggest that initially the effect of deviant parental handling is more discrete than has been supposed.

Although some of the results of this study lead to conclusions which seem obvious at the present date to psychiatrists, psychologists, pediatricians, and other educators in the field of child development, their apparent clarity is probably more the result of hindsight based on ten years' experience than foresight in predicting the effects of a pendulum swing from rigidity to permissiveness in child care. At the formulation of the research in 1946, no actual documentation existed as to the effects of various types of maternal practices on child development. In the ensuing years, two main difficulties arose which probably could not have been anticipated. The first of these was the focus of teaching on the early infancy period, because of professional interest in changing maternity routines, the encouragement of breast feeding, and the emphasis on "ad lib." care during the neonatal period. As a result, although the limits of permissiveness were tacitly implied, they were not explicitly stated. A situation was thereby created in which the mothers were encouraged to handle the child permissively during his infancy, but received less professional assistance in setting limits on such practices as the child grew older. The second difficulty was the rapidity with which the concept of permissiveness was popularized in lay publications. However, here a misinterpretation arose in equating permissiveness with license and in exaggerating the applicability of the new concept. Because of this trend toward overpermissive practices by many mothers, it was possible to document the effects of extremes in maternal care as well as of optimal handling.

Other of the results—particularly the finding of fewer significant relationships in the first year and the discreteness of relationships in the second and third years—may come as a surprise. Whatever explanations may be advanced to account for these results, the fact remains that they are the product of an objectively collected body of data, amassed at the time of origin of the behavior in question. This may in itself account for the disparity between the conclusions of the present study and current theoretical formulations. That is, most of the data which stress the importance of early mother-child interaction and the pervasiveness of problem behavior have been collected from older patients in psychiatric treatment or from a few intensive longitudinal case histories. While not discounting the value and contribution of information obtained from such sources, it is plausible to assume that differences in methods of data collection would lead to emphasizing different aspects of personality development. Where data are collected from patients through retrospective recall, the emphasis is on pathological processes and may include unconscious distortion and falsification of time relationships. Where minute observations of the effect of many individual variables are available, the influence of factors common to all may be obscured. The present study represents one of the first attempts of which the authors are aware to observe and correlate maternal and child interaction in a "normal" population of subjects, with emphasis on the central tendency of the group.

SUMMARY

Fifty records of primiparous mothers were selected from the files of the Yale Rooming-In Project. The data (prenatal interviews; hospital record data; pediatric posthospital follow-up; psychological examination and interview, social worker's home visits, and mother's report by questionnaire at one, two and three years) were analyzed statistically. A rating scale method of analysis was used, with three raters independently rating the mother's child care practices and three the child's behavior. The areas scaled for the mother were; practices in feeding, sleeping, toileting, and socialization, and her adjustment to the maternal role. For the child, the areas scaled were: behavior in feeding, sleeping, toileting, socialization, and evidences of his emotional maladjustment. The scales for maternal child care practice were continua ranging from extreme rigidity at one end to extreme permissiveness at the other, with optimal handling as the mid-point; adjustment to the maternal role was scaled on a continuum ranging from extreme rejection to extreme absorption. All of the scales for the child's behavior were continua ranging from severe problem to absence of problem, with "problem within normal limits" as the mid-point.

The relationships between the consensus judgments of the two groups of raters for mother's practices and child's behavior were analyzed for every area during each of the three years. Tripartite chi-square was first used, dividing the mother's behavior into rigid, optimal, and overpermissive practices, and the child's behavior into degree of problem above or below the median of the group. It was found that the direction of the difference in the mother's scales was of no significance in relation to the degree of problem manifested by the child, and in the final analysis, bipartite chi-square was used, dividing the mother's scales into optimal and deviant practices.

The following statistically significant relationships were found: within the first year, deviant maternal practices in sleep were associated with problem behavior in the child, though no relationships were found between practices and behavior in the areas of feeding, toileting, and socialization. In the second and third years, significant relationships were found between deviant maternal practices and problem behavior in the child in all four areas. Deviant adjustment to the maternal role was found to be related to deviant practices in sleep and socialization during all three years. Similarly, children showing other evidences of emotional maladjustment were found to have a significantly higher proportion of sleep and socialization problems.

Other aspects of the case history (social class, housing conditions, illness in either parent or child, birth of siblings, nursery school attendance, etc.) were analyzed with respect to both maternal practices and child behavior. The only significant relationship found was that between birth of a sibling in the second year and the presence of greater emotional maladjustment in the child. Trends were observed between the mother's social class membership and her type of child care practices.

The significance of this type of cross-sectional statistical study is discussed.

REFERENCES

HOLLINGSHEAD, A. B., & REDLICH, F. C. Social stratification and psychiatric disorders. *Am. Sociol. Rev.*, 1953, *18*, 163–170.

KLATSKIN, ETHELYN H., & JACKSON, EDITH B. Methodology of the Yale Rooming-in Project on parent-child relationship. *Am. J. Orthopsychiatry*, 1955, *25*, 81–108, 373–397.

21

Observation and Parent-Child Interaction

Damaris Pease and Glenn R. Hawkes

The concept of interaction and its analysis in the ongoing situation is the subject of this paper. Previous papers—such as those by Klatskin (20) and Rheingold (11)—have examined the mother and her role and the resulting effects upon the child. Although noting the weaknesses of the method, Pease and Hawkes urge direct observation of the interaction itself, based upon some theoretical position, as the desirable approach. They also stress the fundamental importance of observer training.

Behavior of individuals, as an area of investigation, has intrigued research workers from the beginning of time. The significance of studying parent-child interaction has become increasingly important as research has revealed the effects of interpersonal relationships on the behavior of the growing child and his family.

One of the difficulties in an approach to the investigation of human behavior arises from the segmentation of the individual or group so that only parts or units of behavior are evaluated. Fragmentation of behavior, while contributing to the total body of knowledge, lacks the substance of integration and differentiation which is a vital force in the total functioning of the individual. There is a real need to develop methods for evaluating behavior that include the total impact of the relationship between the uniqueness of the individual and his environment.

Some of the earliest knowledge of human behavior was gained through diary records of infant behavior, usually recorded by one or both parents who were the experimenters. These accounts were valuable in that they gave descriptions of behavior over a period of time, providing clues or suggestions as to the nature and developmental aspects of growth. They were the behavioral records of one child and as such could not be considered representative of a larger sam-

---

From *American Journal of Orthopsychiatry*, 1960, *30*, 453–459. Copyright, the American Orthopsychiatric Association, Inc. Reproduced by permission.

ple or population. The question of biased recording was ever present. The primary purpose of these diary records was to describe behavior rather than to determine its cause.

Because of the subjectiveness of recording observations and the difficulties of evaluating them, observation as a scientific approach to studying human interaction was all but discarded.

With the present increasing emphasis on feelings and attitudes as important aspects of interpersonal relationships the observation method has gained new importance. Direct observation has a unique contribution to make to the study of human behavior, for through it the dynamic qualities of interaction can be studied.

Relationships between people are not static—they grow and mature as the individuals gain insight and understanding of each other. Experiences in successful and unsuccessful interaction with others tend to alter behavior, particularly the behavior of the growing child. Becoming a socialized person is a slow and lengthy process in our society. Frequently the early stages of social development are forgotten by the individual, and while he may honestly try to recall his thoughts, feelings and attitudes, it becomes virtually impossible for him to do so.

On-the-spot observation gives an opportunity for the research worker to evaluate interaction *as it is taking place.* Provided some of the difficulties inherent in the observation method can be eliminated, or at least controlled, observation promises to yield valuable new insights into the behavior of families as they work, play, and grow together.

With the revival of interest in observation as an evaluation device for interaction there arise numerous problems inherent in the method itself. The primary problem in observation is concerned with the quantification of behavior. Units of interaction must be defined and categories of behavior so constructed as to encompass both verbal and nonverbal interaction if a reliable picture of the individual's behavior is to be obtained. A second concern arises with the training of observers and achieving reliability of observed behavior. Finally, the problem of an appropriate statistical model for evaluating the observed units of behavior must be considered.

Sound research technique requires that the research worker have a theoretical basis for the problem he chooses to investigate. Frequently a problem originates from some incidental observations of the researcher—usually some behavior or pattern of behavior that he has seen repeated frequently in a variety of situations. Working with these "clues" he attempts to develop a rational theory to explain their existence. The next step is to test his theory under controlled conditions with a representative sample. It is at this point that he must decide on the design which is most appropriate for the hypotheses being investigated and on the method which will yield the most valid and meaningful data.

If the researcher chooses to use observation as a method for collecting his data, two approaches are open to him. He may take a record of the behavior and at some later date develop categories from which he derives scores, or he may develop appropriate categories for his theoretical framework and record the observed behavior in the categories as the behavior is in progress. Regardless of the approach, the central problem is concerned with developing categories that tend to define the variables to be tested.

Bales (1950) at Harvard has developed a method for categorizing interactions of small groups as they experience face-to-face contacts. Bales' system of mutually exclusive categories, which he has called interaction process analysis, was designed to organize the observation of social interaction. The 12 categories are divided into 3 major areas: social-emotional area, positive; task area, neutral; and social-emotional area, negative. The positive and negative social-emotional areas are related to the affective phase of social interaction; the task area, neutral, refers to the cognitive and conative aspects of interaction. Scoring is done in three dimensions: by categories, by attribution, and by unitization. Attribution is the relationship between the individual who initiates the act and the individual who receives it (who-to-whom relationships). Unitization pertains to the act or unit scored.

Bales' primary concern with interaction process analysis was with the pattern of group problem-solving as it was reflected in the behavior of the individual group members. His observers recorded small group interaction while it was taking place. The problem to be solved was usually a highly structured situation, and the individuals who comprised the group had no previous knowledge of or experience with each other.

It was only a matter of time until other research workers saw in Bales' interaction process analysis the possibilities of investigating interaction among family members. Strodtbeck (1951) studied husband-wife interaction by recording their responses on a tape recorder. The couples were asked to make independent evaluations of shared experiences and then to reconcile any differences in their interpretations of these experiences. The taped protocols were then analyzed, using interaction process analysis, to determine the nature of the decision-making process. Results indicated that joint decisions of husband and wife are related to the amount of participation in the interaction situation and to the power elements in the role differentiation of the cultural organization to which the husband and wife belong.

Using the same technique of recording and analyzing interaction, Strodtbeck (1954) investigated the family as a three-person group including father, mother, and teen-age son. Problems for the group to work on were so constructed as to provide situations in which the father and mother were in agreement, the father and son were in agreement, and the mother and son were in agreement. Thus, one of the family members was always in disagreement.

By organizing Bales' categories into indices, four support types of categories were developed: solidary, contending, dominant, and conflicting. Strodtbeck found that the person who talked the most tended to win the most decisions and the least-speaking person tended to win the least. However, the least-speaking person did not seem to lack support from the other family members. He concluded that decision-making power is associated with high participation in the family, but when the two most active members are in conflict stability is not necessarily low.

Bishop (1951), using a different type of categorization, has studied mother-child interaction as it is related to the social behavior of the child. Preschool children and their mothers were observed during two 30-minute "free-play" periods under controlled conditions. Behavior of both mother and child was recorded at 5-second intervals and the behavior assigned to categories representing facilitative and inhibitory behavior. In general the results tended to indicate that when the mother exhibited behavior considered as directing, interfering, criticizing, and suggesting control, the child responded in an inhibited, noncooperative or negative manner.

An equally important approach to observation of direct interaction does not always involve the construction of a categorical system prior to observation of the actual behavior. Ultimately, of course, some system of quantification must result, but behavior can be recorded objectively without the use of predetermined categories. Barker and Wright's *One Boy's Day* (1951) is a classic example of this type of observation. Raymond was a 7-year-old boy living in Midtown. For 14 hours of one day he was under constant observation by 8 different observers. Barker and Wright made no attempt to analyze the causal factor of Raymond's behavior but submitted this document as a description of what one 7-year-old did from the moment he awoke in the morning until he was tucked in bed at night by his mother.

Gardner and Pease (1958), as part of a larger research program, are in the process of investigating the responses of two-year-old children to certain controlled stress situations. One such situation is called the Stranger Relationship Test. When the investigators leave the room the child's mother tells him that she must leave for a minute but that she will be right back. The child is then left in the room alone. In 15 seconds a woman, previously unknown to the child, enters the room and sits down for 45 seconds. At the end of this time she asks the child if he would like to go and find his mother. They then leave the room and find the mother, who is waiting outside.

During this 60-second interval the experimenters jointly record the behavior of the mother and child, the child alone in the room, and the child with the stranger. A one-way vision mirror is used, and as one observer reports the behavior the other records. Frequent checks by the recorder are made on the reported behavior.

If this type of observation is to yield meaningful data certain rules or insights into objective recording need to be understood by the observers. Since individuals bring to any situation a multitude of personal experiences by which they interpret present behavior, observers need to be aware of the differences between this personal interpretation and the observable facts. However, if only behavior which might be recognized by everyone is recorded and the impressions or interpretations of the observer are ignored, much valuable information about interaction is lost.

One method for retaining both types of information is to recognize the observer's bias and record it. The observational record can be divided into two parts: one pertaining to facts and one to interpretation. As the observation is recorded an attempt is made to describe behavior that is actually seen (facts) and then to interpret this behavior. For example, an observer records the behavior of four-year-old David as he greets his mother at nursery school. The observer sees that David and his mother hug each other and records: "David and his mother are very affectionate." This gives little if any information about what actually happened between David and his mother but rather is an interpretation by the observer of what was seen. Objective recording would show: "David went over to his mother and put his arms around her. He puts his hands on either side of her face and gives her a kiss. David's mother hugs him and takes his hand as they go out the gate." Interpretation by the observer might include: "David likes to kiss his mother. He is very affectionate." This type of recording requires skill in being aware of the activity under observation and of value judgments inherent in certain words and phrases.

Regardless of the type of observational approach, i.e., observation based on a previously determined categorical scheme or on a diary recording, observers need to be trained so that they understand the function of the categories or the level of objectiveness of the diary record. The importance of observer reliability cannot be overemphasized, for their scores or recordings represent the raw data by which specific hypotheses will ultimately be tested.

One method of training observers which seems to hold promise is the use of motion pictures. The particular advantage is that the behavior can be viewed many times without altering the interaction situation. In a study by Brady (1955) in which she investigated the interaction of brothers, an attempt was made to establish observer reliability by first training the observers in the use of interaction process analysis through a filmed sequence.

For a little over a month observers met three times a week to score the interaction seen in the motion picture. At the end of this training period it was found that observer reliability, deterined by chi square, when tested on a film not previously seen resulted in probability levels varying from .2 to .7. The pair of observers having the highest reliability were then trained in live sessions and probability levels between .7 and .9 were obtained. However, when these observers continued to score live sessions the probability levels were variable and inconsistent. Brady felt that some of the inconsistency of performance between observers might be accounted for by the fact that additional skill in scoring was being acquired at a different rate of competency by each observer.

Howell (1956), using Bales' technique, studied the interaction between students and preschool children as a measure of the effectiveness in attitude changes of college students enrolled in a basic child development course. She used the pair of observers previously employed by Brady. Although live training sessions were initiated, Howell experienced difficulty in establishing consistent levels of reliability for the observed behavior.

If there is justification in drawing tentative conclusions from these studies it might be suggested that when precise and definite categories of interaction are used in live sessions, training of observers needs to take place over a relatively long period of time. It is interesting to note that in both studies when the observer reliability was calculated using the broader areas of positive, neutral, and negative interaction rather than the 12 more specific categories, probability levels were consistently in the neighborhood of .7 and 8. This suggests the question of finiteness vs. grossness in developing a categorical scheme.

The problem of the effect of the observer or observers on the interaction situation deserves some consideration. There can be little question but that the intrusion of the observer alters the interaction situation. Recognizing this, it is possible for the research worker to minimize this effect by using observation screens or booths or by so constructing the interaction task as to take advantage of the additional stress produced by the presence of the observer.

In some instances the observers can be so much a part of the environment that the subjects show little effect from their presence. Barker and Wright felt this was the case in their account of Raymond and his eight observers.

An example of taking advantage of the presence of an observer can be illustrated by a study in progress at Iowa State University (Pease & Gardner, 1958). As part of a larger study mothers bring their year-old children to the Child Development Research Laboratories to undergo certain psychometric tests and developmental and anthropometric measurements. The test situation varies in length but usually takes an hour and is terminated by the anthropometric measurements, which most year-old children resist strenuously. Immediately following the physical measurements the mother dresses the child on a table approximately the size of a card table. She is instructed to keep the child on the table and to go about the dressing procedure as she normally does. During this time a motion picture record is made of the interaction between the mother and her child. There is little doubt that the mother reacts to the presence of the camera and behaves differently under this artificial stress situation than she does in the privacy of her home. However, the purpose of the recording is to observe mother-child interaction under stress. The camera and operator act as additional pressure to the situation and are a consistent force from test to test.

Filming of parent-child interaction is a promising type of recording that eliminates many of the disadvantages of live observation. It provides an opportunity for the experimenters to review the interaction without alteration of the situation and gives a documentary record of behavior. It allows analysis at a later date without loss of data. The interaction behavior may be looked at in terms of a variety of variables without the contamination of observer interpretation.

Filmed sequences have their limitations, for the camera can record only that behavior which it sees. All the activity on the periphery is lost, some of which can be extremely meaningful. The effect of the actual filming on the subjects cannot be eliminated but it can be minimized. An experimental room can be equipped so that the camera and operator are housed in an area not observably a part of the room, and while the subjects know they are being filmed the actual filming is not obvious to them.

It seems appropriate to end this discussion by suggesting some types of information pertaining to parent-child interaction that can be appropriately investigated through the observation method. Studies of parent-child interaction under stress conditions could yield information as to the stability of the parent-child relationship, and the patterns of behavior employed in meeting such situations. Decision making, as a variable of investigation, might reveal information as to how parent and child differentiate their roles in the family, how the dependent-independent relationship of the child is established, and how the strength of the parental or child role is determined. Guidance is an important aspect of family responsibility, and knowledge of how parents pass on family and cultural values could give insight into the developmental pattern of the socialization of the child.

It would seem that research centered around parent-child interaction would give fruitful new insights into the development of personality and the structure of family living and would be of value not only to clinicians, research workers, and social caseworkers, but to parents and educators as well.

REFERENCES

BALES, R. F. *Interaction process analysis.* Cambridge, Mass.: Addison-Wesley, 1950.

BARKER, R. G., & WRIGHT, H.F. *One boy's day.* New York: Harper, 1951.

BISHOP, BARBARA M. Mother-child interaction and the social behavior of children. *Psychol. Monogr.*, 1951, *65*, No. 11.

BRADY, MARY ROSE. Interaction of brothers in an experimental play situation. Unpublished masters thesis, Iowa State University, 1955.

GARDNER, D. B., & PEASE, D. The use of situational tests with preschool children. *J. nursery Educ.*, 1958, *14*, 18–20.

HOWELL, GRACE. Changes in student-child interaction during an elementary child development course. Unpublished masters thesis, Iowa State University, 1956.

PEASE, D., & GARDNER, D. B. Research on the effects of noncontinuous mothering. *Child Develpm.*, 1958, *29*, 141–148.

STRODTBECK, F. L. Husband-wife interaction over revealed differences. *Am. sociol. Rev.*, 1951, *16*, 468–473.

STRODTBECK, F. L. The family as a three-person group. *Am. sociol. Rev.*, 1954, *19*, 23–29.

22

Adjustment of Problem and Nonproblem Children to Specific Situations: A Study in Role Theory

Harry H. L. Kitano

Kitano's study utilizes observation in the manner suggested by Pease and Hawkes: trained observers categorized the behavior of youngsters in specific interaction situations; and the situations, ratings, and children were selected to test an aspect of theory. His conclusions suggest that children behave in ways appropriate to a given situation and that their behavior cannot be evaluated apart from that situation. For this reason, the concept of a "problem" child needs to be re-examined.

From *Child Development*, 1962, *33*: 229–233. Reprinted with permission of the Society for Research in Child Development.

One difficulty in dealing with "problem children" lies in the selection or identification of those who constitute problems. For example, emphasis may be put on active or on passive maladjustment. As studies by Wickman (1928) and Beilin (1959) show, teachers are most aware of overt and aggressive behavior as a problem, while mental hygienists are more concerned with the type of emotional trouble which is often associated with withdrawn and other nonsocial behavior. There has been some convergence between the views of teachers and mental hygienists since the early Wickman studies, but complete congruence is unlikely to be achieved, nor perhaps is it desirable. Each profession naturally perceives the problem from its own point of view.

It is believed that role theory as developed by Parsons and Shils (1951) and Sarbin (1954) offers a sound theoretical basis for formulations concerning culture, society, and personality. In broad perspective this theory regards human conduct as a product of the interaction of self and role. The purpose of this study was to use concepts derived from role theory in order to test a hypothesis concerning problem behavior originally advanced by Cottrell (1942) which deals with specificity of role expectations. Cottrell writes: "The degree of adjustment to roles which a society assigns to its age-sex categories varies directly with the clarity with which such roles are defined" (1942, p. 618). If this is so, we would expect to find that children are not problems per se or in isolation but that they become or are perceived as problems in relation to a dimension of expectations.

PROCEDURE

The subjects for the study were 46 children, 22 boys and 24 girls, enrolled in two day care centers of a child care program. Most of the children were from one-parent homes with mothers who were employed full time. The children spent from eight to ten hours a day in the centers. The ages of the children were from 2½ to 5. Fourteen of these children, 8 boys and 6 girls, were considered by the staffs of the centers to be serious problems or to be seriously disturbed. Therefore, 32 children constituted the normal group and 14 children were in the problem group. . . .

Several typical nursery school situations such as lunch, waking-up time after nap, dressing, going home, and free play were chosen for observation. The child care teachers, eight respondents, were asked to write descriptions of desirable and undesirable child behavior in each of these situations. The answers were analyzed, and in the final study data was gathered only for those situations where the expectations were similar. It is of interest to note that the descriptions of what constituted favorable and unfavorable behavior in situations such as free play (e.g.: "the child should play in a group"; "the child should play independently") varied so widely that identical behaviors would be classified as "normal" and as "problem" from the point of view of different personnel.

A 7-point rating scale was then prepared for each of the remaining four situations: lunch, waking-up time after nap, dressing, and going home. An example follows:

SITUATION: DRESSING

1	2	3	4	5	6	7
dresses self independently			needs some help and direction			needs full attention

Two graduate students in social welfare each observed and rated the children in both schools several times. Independent prerating and subsequent checking between the two observers and the use of the behavior rating scales made for reliable observations. The raters were never told which children constituted the "problem" and "normal" groups.

RESULTS

The mean scores of both groups of children are given in Table 1. (The total rating of each child was the mean score obtained from one, two, or three observations: the number of observations of individual children varied because of absence at the times of rating.) In the case of "going-home time" the behavior of both parent and child was rated.

In all of the situations, the behavior of both groups of children and parents fell at the favorable end of the scale. In other words, the behavior of this sample was congruent with teacher expectations.

Total mean score for all situations was 2.3 for the "problem" children and 2.2 for the "normals." A *t* test of group means indicated that the difference was nonsignificant. Inspection of

Table 1: *Mean scores and standard deviations of problem and normal children in selected situations*

Situation	Problem Group $N = 14$		Normal Group $N = 32$	
	Mean	SD	Mean	SD
Lunch	1.4	.6	2.1	1.0
Wake-up	1.5	.6	2.0	1.4
Dressing	3.4	1.8	2.4	1.7
Going home:				
Child	2.6	1.7	2.0	1.3
Parent	2.4	1.3	2.2	1.3
Total	2.3	1.5	2.2	1.3

mean scores of the groups in the specific situations also indicated nonsignificant differences leading to the acceptance of the null hypothesis of no difference between the two groups in behavior in these situations.

The raters were also asked to comment on each of the situations from a more subjective point of view. A typical comment was: "I have never seen such a well-behaved bunch of kids." This was in direct contrast to the observed behavior of these same children in other than the specified situations.

DISCUSSION

Gross, Mason, and McEachern (1958), in analyzing role behavior, indicate the fruitfulness of looking at role consensus as a social science variable. "That the members of a social system, whether a dyad or a total society, must agree themselves to some extent on values or expectations is a matter of definition. The point we have been trying to underscore is that the degree of consensus on expectations associated with position is an empirical variable whose theoretical possibilities until recently have remained relatively untapped" (Gross, Mason, & McEachern, 1958, p. 43).

The staff of the Child Care Centers comprises a social system, and an analysis of consensus in terms of role expectations towards children in specific situations was one variable studied here. Initial attempts at obtaining role consensus as to what constituted "normal" and "problem" children on a more global level were not successful. The behavior of children in relation to specific situations was therefore selected as an object of study. It was found that only in the case of four specific situations was there agreement as to what constituted "normal" and "problem" behavior.

The other variable for this study was the actual observable behavior of the children in each of the specified situations. The children were divided into a "problem" and a "normal" group on the basis of interviews with the staffs. The hypothesis was that the "normal" group would show a higher degree of adjustment (as measured by the rating scale) to each of the situations than the "problem" group.

Analysis of the data disproved the hypothesis. There was no significant between-group difference in total score or in scores for any of the four specific situations. What was interesting was the high degree of observed adjustment of all children to each of the situations. It appears that, if role prescribers (in this case teachers) will hold expectations for behavior which are similar and reasonable, then there will be little variation in the behavior exhibited by children in these situations. This was true even of children who were considered serious problems.

Where role expectations were clear, there was little room for confusion on the part of the children. Conversely, in situations where there was little consensus, there was much pupil confusion, and it was obvious that teachers encouraged or discouraged behaviors on an individual basis—and at times different teachers encouraged completely opposed behaviors. This is, of course, not to be considered as an "unhealthy" situation, but it points up the fact that children who are diagnosed as problems appear to be most maladjusted in these unstructured situations.

What constitutes a problem child can then be analyzed from the point of view of expectations and behavior. A child who "hits" or "withdraws" cannot be regarded as a problem until (a) the situation and (b) the expectations of the observers have been analyzed. This is an unstated but obvious factor in all criteria for establishing "problems." For example, in most standardized psychological tests, expectations are associated with the test norms; discrepancies between the individual taking the test and the norms are usually seen as an indication of disturbance. Some of the conflicts between different professions regarding the diagnosis of a problem might be resolved by an analysis of the expectations common to various roles and positions, as well as by the more traditional method of concentrating on the "client."

Today, many clinicians restrict their interest to the personality of the problem child. The findings of this study suggest that other variables, such as the demands and expectations placed upon the child, should also be explored and analyzed. Only through this broader approach can we reach a fuller understanding of children's behavior.

SUMMARY AND CONCLUSIONS

This study explored a hypothesis drawn from concepts derived from role theory. The findings indicate that, in situations where role expectations are clear, where there is consensus, and where role prescriptions are readily enforceable, there is a high degree of adjustment on the part of both "normal" and "problem" children.

It also appears that the diagnosis of a "problem child" is based on other evidence than that supplied by his or her behavior in this type of specific situation. It is suggested that an analysis of the total situation of the child, including the various demands and expectations placed on him, should be part of any definition of what constitutes a "problem child."

REFERENCES

BEILIN, H. Teachers' and clinicians' attitudes toward the behavior problems of children: a reappraisal. *Child Develpm.*, 1959, *30*, 9–25.

COTTRELL, L. S. JR. The adjustment of the individual to his age and sex roles. *Amer. sociol. Rev.*, 1942, *7*, 618–625.

GROSS, N., MASON, W. S., & MCEACHERN, A. W. *Explorations in role analysis.* New York: Wiley, 1958.

PARSONS, T., & SHILS, E. A. *Toward a theory of action.* Cambridge, Mass.: Harvard University Press, 1951.

SARBIN, T. R. Role theory. In G. Lindzey (Ed.), *Handbook of social psychology.* Vol. V. Cambridge, Mass.: Addison-Wesley, 1954. Pp. 223–258.

WICKMAN, E. K. *Children's behavior and teachers' attitudes.* New York: Commonwealth Fund, 1928.

COGNITIVE DEVELOPMENT

23

The Development of Perception in Children

M. D. Vernon

Writing for teachers in the English schools, Vernon offers practical applications and cautions growing out of the evidence available from the literature on perceptual, language, and number development. Although her point of view is maturational, she clearly recognizes the role of experience in development. The journal in which this paper appeared, Educational Research, *is a prime source of material for the student who seeks readable syntheses of research with application to the problems of education.*

INTRODUCTION

It is often difficult for adults to realize that children, even when they have learnt to run about and to speak, do not perceive the world around them in the same way as do adults. Children are so surprisingly quick to notice and comment on everything that goes on; yet at the same time the things they perceive, and the man-

---

From *Educational Research*, 1960, *3*, 2–11.

ner in which they perceive them, may differ quite considerably from the things we notice. It is perhaps particularly important for teachers of young children to recognize these facts clearly, and to allow for them in their teaching, since misunderstandings may occur when children seem wilfully to overlook or to make mistakes in perceiving what is shown them. Nor should it be forgotten that the ability to perceive in an adult manner develops probably more through maturation than through learning. True, the child requires certain experiences to expand and refine his perceptual capacities; but no amount of experience, nor even of direct teaching, can force him to proceed beyond the stage which he has reached through natural maturation.

It is often not all easy to determine the stage of maturation. It does not depend solely on chronological age, but neither is it entirely a function of intelligence, particularly of intelligence as measured by an intelligence test. Emotional maturity is also of importance; and so is the experience which the child has gained from his physical and social environment. When in the subsequent discussion it is stated that children of a certain age in years can perform such-and-such a task, it must be understood that this age is merely an average. Many children will be either older or younger when they can first perform it.

PERCEPTUAL DEVELOPMENT IN EARLY CHILDHOOD

Before we examine the gradually developing perceptual abilities of the child when he first goes to school, it is important to consider the outlook of the pre-school child in perceiving the world around him. His main concern is with the objects and events he encounters in his ordinary everyday life, particularly those which are useful and interesting to him. He needs to know what things are; what they do; what he can do with them. Though in early infancy he has no understanding of the nature of his environment, nor what objects are like, he can discover a good deal, as Piaget (1955) has shown, by examining closely everything that comes within his reach, putting it in his mouth, manipulating it and exploring it with his fingers. In tnis way during his first year he comes to realize that objects are stable, solid, resistant to touch; and that they have a shape which remains the same although its visual appearance varies as he turns the object around in his hands. He knows also that objects retain their identity although their apparent size varies as they approach and recede from him; and that they continue to exist while they are out of sight. Before he realizes this, he tends to think that a hidden object has altogether disappeared, and seems quite surprised when it reappears. The realization of the continued existence of objects out of sight is very important to the development of understanding, since it enables the child not only to ask for things not perceived at the moment, but later to think and talk about them. . . .

During his first year the child also begins to learn something about cause and effect. He is extremely interested in pushing and pulling things and making them move about or rattle. But for some while he does not realize that it is necessary to touch them in order to do this, and sometimes he tries to "magic" them into moving by waving his hands at them from a distance. During his second year he also comes to understand that a moving object must touch another object to make it move in turn. Thus, even at this period he is extremely interested in finding out, as far as he is able, how and why things happen.

In his second year the child begins to use language to help him to understand and come to terms with his surroundings. Earlier, during his first year, vocalization is used primarily as a means of getting what he wants, crying when he is hungry or needs attention. Even his first words, "Mama," "go," "up," are really "word sentences," expressing his desires and needs. Listening to, smiling at and babbling in response to human voices seems to be carried on in order to establish social contact with the people around him. Soon he finds that he can get what he wants more easily by asking people for it than by crying or grabbing. Also he often seems to enjoy babbling and repeating his early words as a form of vocal play, just as he enjoys waving his arms and legs and playing with his fingers and toes.

The child's first words are most often names of objects, sometimes spoken in order to obtain these objects, but sometimes as a means of identifying an object and knowing what it is. It seems that the child establishes this identity in his mind by giving the object a name. At first the names are used for certain particular objects, but after a time for classes of objects. Thus, Lewis's son at about 9 months would look and reach for a small white ball when someone asked him, "Where's ballie?" (Lewis, 1936). But it was not until he was over a year old that he responded in the same way to a large colored ball. However, as the child begins to encounter more and more objects, he comes to perceive that certain of them are alike. He hears adults calling these similar objects by the same name, and so he begins also to class them together and give them a common name. Thus, balls may be big or small, hard or soft, white or colored; but they all have the same name. This classificatory procedure is extremely useful to the child, since instead of remembering the nature and identity of each separate object, he can, once he has classified them, perceive and recognize them as a group of things of similar behaviour and use (Vernon, 1952). Thus, balls are things which can be thrown and caught, and which bounce. When the child encounters a new object, he immediately tries to classify it and to know what sort of a thing it is. If he cannot do this, he asks, "What's that?" Thus, we find that "What" questions begin towards the end of the second year, and are exceedingly frequent for the next two years.

Of course, at first the child makes many mistakes in classification and naming. A single name may be used for a number of objects roughly similar in appearance to which we give quite different names. . . . Thus, the child develops his own schemes of classification before he learns to name and use them in the way that we do.

As the child begins to know what things are, he becomes increasingly interested in what they do and what he can do with them and use them for. Thus, Valentine (1942) noted that his two-year-old daughter was constantly asking, "What this for?" A rather older child of five, when asked what various things were, described them in terms of use and function: "What is a hen?" "Something that lays eggs for you." The child also needs to acquire an understanding of how and why things happen, what people and things do and why. This is reflected in his questions as to how and why, which become increasingly frequent in the fourth and fifth years. Indeed, it has been calculated that at this period questions form 10–15 per cent of everything the child says (Fisher, 1934). Even what appear to be statements of fact are often made with the intention of inviting confirmation or contradiction.

But in the process of classifying objects according to their appearance and use, there is naturally a tendency to overlook details which are not important for identification. The child obtains a broad general impression of the object as a whole, but pays little attention to the specific qualities of particular objects. Neither does he notice these qualities, shape, color, size, weight, in themselves, as apart from the objects to which they belong. The purpose of Montessori teaching was to encourage children to abstract such qualities and scale them in order. With specially prepared material, in which only one quality varies at a time, this is easier for the child than it is with ordinary objects which, as we have noted, the child regards mainly from the point of view of general appearance and use. We shall see later that these processes of abstraction and generalization constitute a real difficulty for the child when he begins to be taught formal school subjects.

When he first comes to school, the child needs to discover the nature and identity of the features of his new environment. He wants to know what things are, how and why events happen. Thus, he requires explanation rather than instruction. Still more, perhaps, he needs the opportunity, as he did in infancy, to work things out for himself by exploration, experimentation, manipulation and construction. Susan Isaacs, in her book *Intellectual Growth in Young Children* (1930), gives numerous descriptions of the experimental play carried out by the children at the Malting House School at Cambridge, with sand, water, building blocks, water pipes, gas burners, etc.; and how they weighed, measured, melted things, dissolved them in water, and so on. These experiments, the discussions of them and the explanations given by the teachers, enabled the children to learn and understand much about the nature of materials and the causes of physical phenomena. Their natural intense interest in how and why was furthered; and their powers of reasoning developed and were reinforced.

DEVELOPMENT OF VISUAL PERCEPTION

We see, then, that the primary need of children is to learn to understand and deal with "real" things and events. There is considerable evidence to show that it is some while before they have any desire or capacity to perceive things which may be educationally important, such as details of shape, pattern, arrangement and number. It is true that children can learn at the age of two years or even younger to pick out a simple shape, such as a square, a circle and a triangle, from among others if their choice is rewarded by a piece of chocolate. And by four years a child of average intelligence can match eight out of the ten outline shapes used in one

Fig. 1

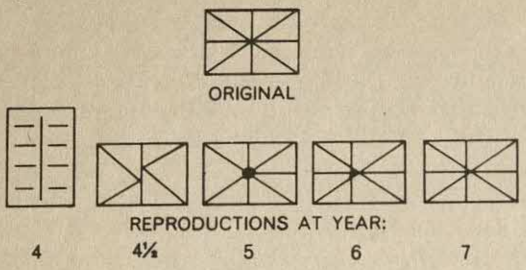
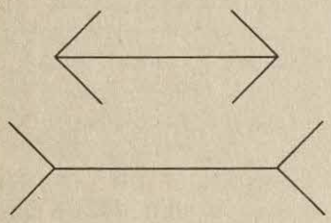


Fig. 2



of the Terman-Merrill test items (Terman and Merrill, 1937). But it is more doubtful if they can remember these shapes. Piaget and Inhelder (1956) showed that a circle and a square could not be copied correctly before about four years of age, a triangle before about five and a diamond before six. With complicated shapes made up of several parts or containing interior detail, accurate perception is slower to develop. This is perhaps best illustrated by Gesell and Ames (1946) in the reproductions made by children of various ages of a figure like a Union Jack (see Figure 1). Similar mistakes in the reproduction of complex forms have been noted by [other researchers]. They all indicate that children under six or seven years do not grasp the relationship between the parts of a complex figure. They may perceive the outline, or some of the parts separately, but not the manner in which these are fitted together. The ability to do this may depend as much upon the understanding of relationships as upon immediate perception of shape. Thus, when children begin to learn to read, they may notice the length of a word, and certain letters in it. But they may not perceive or remember correctly the exact shapes of any letter, nor all the letters in the word, nor the relationship or order in which the letters occur.

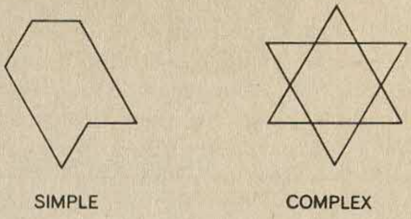
There is another difficulty frequently encountered in teaching children to perceive accurately, namely, that they tend to see things as part of a total situation from which they cannot be isolated. An instance of this is given by Piaget (1955) of a child of 1½ years who used the name "papa" only for her father sitting in his study; when she saw him elsewhere she did not associate this name with him. In older children this phenomenon appears in an inability to differentiate shapes from their background, and to perceive shape as such independently of its surroundings. Thus, children are particularly susceptible to visual illusions such as that shown in Figure 2, in which the capacity to perceive that the two horizontal lines are of equal length depends upon isolating them from the attached arrowheads (Walters, 1942). Piaget (1960) has also made a large number of studies of children's perceptions of visual illusions and concluded that they do not know how to direct attention towards the significant features of the figures and ignore the less relevant. If the children's eye movements are recorded, it is found that the eyes of the younger children wander about all over the figures; whereas those of older children concentrate on particular points of importance.

That children find difficulty in analysing complex shapes and picking out certain parts of them is also shown by their reactions to figures of the type designed by Gottschaldt (1926). He presented a number of simple figures until they were thoroughly familiar. Afterwards he showed complex figures each of which contained one of the simpler ones embedded in it, and found that adults had great difficulty in perceiving the simple figures embedded in the complex ones. Ghent (1956) presented a simple and a complex figure together (such as those shown in Figure 3) and asked children to trace out the simple figure in the complex one. No children could do this before six years, and many not even at eight years. But Witkin (1960) has recently brought evidence to show that the ability to re-structure what is presented by analysing it and extracting particular parts is a function not only of age and intelligence, but also of certain qualities of personality. He considers that children who are in themselves self-reliant and able to act on their own initiative are more capable of such tasks than are children who tend to relay passively on external circumstances or depend on others.

There is evidence to show that the ability to perceive shape in general more accurately is not greatly improved by giving practice in perceiving particular shapes. Nor is the ability to perceive letters and words in reading much affected. Thus, Goins (1958) gave six-year-old children extensive training in perceiving figures and numbers exposed for a short period of time only. They became more skilful in perceiving this particular material; but their achievement in reading improved no more than did that of similar children who were given no such training. Again, as Gates (1922) and Kendall (1948) have shown, there appears to be little or no relationship between ability to perceive shapes correctly and to learn to read. This at first sight seems surprising. We can only suppose that by the time children are sufficiently mature to learn to read, they have developed an ability to perceive shapes visually which is adequate for the purpose of reading; whereas less mature children do not improve sufficiently in perception as the result of special training for there to be any effect on their reading capacities.

There is another form of perceptual ability which may be of considerable indirect importance in learning to read, namely, the perception and understanding of pictures. These also seem to depend to a considerable extent on maturation, though naturally they do not develop unless the children are given some experience with pictures. Binet (1908) pointed out that this understanding developed only gradually; and his observations have been confirmed by later experiments (Vernon, 1940). Up to the age of 4–5 years, children, though they can identify familiar objects in pictures, will enumerate them one after the other but will not relate them together in any way. At about 6–7 years they will begin to give simple descriptions of the pictures; but it is not until they are 10–11 years that they can "interpret" what is happening, what the people are doing, and so on. Thus, children of 11 years are supposed to be able to interpret the picture in the Terman-Merrill test (1937) of the telegraph boy whose bicycle has broken down and who is signalling to a car to stop and give him a lift. Naturally, the age of interpretation varies somewhat with the obviousness and familiarity of the incidents depicted. But before they are able to interpret, children may not even pick out the salient features of the picture or the principal actors in the scene depicted, but may notice instead comparatively irrelevant features. Thus, in a picture used by the author (Vernon, 1940) of a man stopping a runaway horse and cart, the man himself was sometimes omitted from their descriptions, but people looking on were included. This behaviour of course parallels that which appears in children's perceptions of complex forms. The inability to extract the significant features of pictures may be of some importance if children are taught to read words describing the actions of people shown in pictures.

Fig. 3



LANGUAGE DEVELOPMENT

We have already noted that in their early language development children learn first to express their wishes and then to name objects and ask questions about them. At first, the phonetic pattern of children's speech is so unlike that of adults that few except their mothers are able to understand them. But Morley and Court (1957) stated that among a sample of over 100 children they tested, about 70 per cent were intelligible to strangers by the age of two years, though they tended to relapse into unintelligibility when they were excited or upset. Sheridan (1948), however, found that the phonetic pattern of speech of five-year-old children was frequently defective and sometimes so much so as to make their speech unintelligible outside their families. Therefore, when they first come to school such children must have considerable difficulty in making themselves understood.

Children's speech also differs grammatically and syntactically from that of adults. Towards the end of the second year they speak in short 2–3 word sentences consisting mainly of nouns and verbs, prepositions and conjunctions being omitted (McCarthy, 1930). Sentences increase in length from about 4 words each at 3 years to 7½ words at 8 years, according to Templin (1957). Grammatical errors, at first frequent, decrease in number; sentences become more complex in structure and subordinate clauses begin to appear.

It is generally agreed that children understand speech before they speak themselves. Even in early infancy children can often understand and respond to the emotions of adults speaking to them; they comprehend the intonational pattern of adult speech, which reflects the emotions of the speaker, before the phonetic pattern. Then also the adult helps the child to understand words by gesturing, pointing to objects named, and so on. It is difficult to judge how soon and to what extent children hear the phonetic patterns of single words and understand their meanings. . . .

The speech and understanding of children clearly varies greatly according to the manner in which they have been brought up. Those whose parents themselves speak well, and frequently talk to their children, will learn much faster than those brought up by parents with phonetically and grammatically incorrect speech and limited vocabulary. Several studies have shown that in middle-class children both articulation and language structure are more advanced than in working-class children (McCarthy, 1930; Davis, 1937); and, indeed, the differences may increase as the children's ages increase. Bernstein (1958) has put forward the thesis that children in working class homes grow up to speak and understand mainly what may be called "public language": short, crude and obvious statements about immediate feelings, wishes and actions. Middle-class children, especially those from educated homes, become accustomed to hearing "formal language": more complex and more logical language conveying the finer shades of meaning and expressing feelings and wishes more subtly and indirectly. So also objects are not taken simply at their face value, but become a point of departure for enquiry and for interpretation of more recondite meaning and significance. These constitute the language and outlook of the school. Not surprisingly, therefore, the middle-class child is more ready than the working class child to accept and profit by his school education. In fact, the latter may resent and react against school teaching. In support of this thesis Bernstein found that working-class boys aged 15–18 of normal intelligence often gave subnormal performances on a vocabulary test; and, indeed, the higher the non-verbal I.Q., the greater the difference between it and the vocabulary test score. Although these were older children, the same type of difficulty and the same discrepancy may well appear in younger children.

But also close personal contact with adults is immensely important in speech development. Kellmer Pringle (1960) found that children brought up in institutions, without close personal contacts with adults, were particularly backward in their ability to understand and to express themselves in speech. Their linguistic development was slower than their intellectual development. Speech develops slowly in children who are mainly in contact with other children. McCarthy (1930) found that children who associated mainly with adults were more advanced than those who associated chiefly with other children; and Davis (1937) showed that in only children speech was more advanced than in others. Retardation may be particularly marked in twins (Davis, 1937). The Russian psychologists, Luria and Yudovich (1959), described a pair of identical twins, aged 5 years, whose mother was retarded in speech, and who, when they first came to school, were so inarticulate as to be incapable of playing with the other children. But they could communicate with each other by noises and gestures. It is often alleged that the children of over-indulgent parents are liable to be backward in speech, because the parents give them everything they want without requiring them to ask for it intelligibly. We do not know if such children are equally backward in understanding speech; but since speaking and understanding speech are closely related to each other in childhood, it seems possible that they are.

Thus, we see that unless in early childhood they have been helped and encouraged by adults to speak and understand clearly, children have little capacity for, or interest in, listening to what the teacher says. Neither can they express their ideas or talk conversationally.

But in order to learn to read children must not only be able to hear and understand words spoken to them, the meaning of which can be understood from the context of the sentence in which they occur. The child must also hear the exact phonetic pattern of the word, and, indeed, be able to analyse this into its constituent sounds, the sounds of the letters or letter groups. Such a process of analysis is probably even more difficult than the analysis of the printed word shape into its constituent letter shapes. . . . The child hears sentences as wholes conveying instructions or statements. He is not clearly aware of the words as isolated and independent entities. Still less is he aware of the particular letter sounds within the word. Yet before he can read properly he must be able first to isolate the letter sounds from the word sound, and then associate the independent letter sounds with their shapes in the printed word. Unfortunately, in the English language the vowels have many different sounds, and the child may be unable to tell which is the correct one until he has sounded the whole word. Furthermore, the child may be accustomed to hear and speak vowel sounds with a local accent which distorts them until they are quite unlike those of the teacher.

DEVELOPMENT OF NUMBER CONCEPTS

It appears, therefore, that in many of the school tasks which confront the child he cannot go far as long as he is limited to the use of his spontaneous perceptions. He must use his intelligence to refine these and improve their accuracy. And he must learn to break up what is immediately perceived, to analyse it into its constituent parts and use these independently of the whole pattern of sight or sound in which they are presented.

The ability to analyse and abstract particular features from what is perceived is particularly important in beginning number work. Piaget (1952) has made a number of studies which demonstrate the inability of children below the age of five or six to understand the abstract notions of number, quantity and volume. Thus, if they are shown a number of blocks scattered over a wide area, they say that there are more of these than when the blocks are placed close together. Again, if they are shown water in a wide-necked vessel which is then poured into a narrow-necked vessel, they say that there is now more water than before, because it rises higher in the neck of the vessel. They cannot conceive of the number or volume remaining constant, whatever the manner in which objects or substances are distributed; and do not realize that these are qualities which are independent of the objects or substances quantified and the setting in which they are placed.

In a recent pamphlet Nathan Isaacs (1960) has described some of Piaget's observations of children's concepts of number and quantity, and has pointed out how impossible it is for children to do number work satisfactorily until they really understand that number is an abstract quality, independent of the particular objects numbered. If the children are taught number work before they are quite clear about this, names of numbers and computation processes become a kind of verbal ritual which they learn by heart without understanding in the least what it means. Fortunately, many children seem to acquire this understanding spontaneously as their powers of reasoning mature. But others may require prolonged exercises in estimating numbers of objects in different settings.

CONCLUSIONS

We may conclude that both in vision and in hearing perceptual development is active rather than passive. The child does not rest content with a passive reception of the information from the outside world which impinges on his senses. Rather, because of his need to find out, to understand, to get what pleases him, he sets out to explore and to investigate for himself; and to induce adults to give him the objects and the information which he requires. All the time he uses his powers of reasoning to seek knowledge about how, why and wherefore. At first he can reason only by action—by doing things and discovering what happens. But verbal reasoning develops as an accompaniment to such activities. Perceptual and reasoning abilities improve through natural maturation: but they also require the opportunity for exercise, and they require encouragement and help from adults. Children are able to profit from formal teaching in school only when they have reached the necessary stage of maturation, and if and when this teaching appeals to their natural interests.

REFERENCES

BERNSTEIN, B. Some sociological determinants of perception. *Brit. J. Sociol.*, 1958, *9*, 159.

BINET, A., & SIMON, T. Le développement de l'intelligence chez les enfants. *Ann. Psychol.*, 1908, *14*, 1.

DAVIS, E. A. *The development of linguistic skill in twins, singletons with siblings, and only children from age 5 to 10 years.* Minneapolis: University of Minnesota Press, 1937.

FISHER, M. S. Language patterns of pre-school children. *Child Develpm. Monogr.*, 1934, No. 15.

GATES, A. I. The psychology of reading and spelling. *Teachers' Coll. Contrib. Educ.*, 1922, No. 129.

GESELL, A., & AMES, L. B. The development of directionality in drawing. *J. genet. Psychol.*, 1946, *68*, 45.

GHENT, L. Perception of overlapping and embedded figures by children of different ages. *Amer. J. Psychol.*, 1956, *69*, 575.

GOINS, J. T. Visual perceptual abilities and early reading progress. *Suppl. Educ. Monogr.*, 1958, No. 87.

GOTTSCHALDT, K. Uber den Einfluss der Ehrfahrung auf dic Wahrnehmung von Figuren. *Psychol. Forsch.*, 1926, *8*, 261.

ISAACS, N. *New light on children's ideas of number.* London: Educational Supply Association, 1960.

ISAACS, S. *Intellectual growth in young children.* London: Routledge, 1930.

KELLMER PRINGLE, M. L. Learning difficulties of deprived children. *Proc. XVth Int. Congr. Psychol.*, Brussels, 1957. [1960]

KENDALL, B. A note on the relation of retardation in reading to performance on a memory for designs test. *J. educ. Psychol.*, 1948, *39*, 370.

LEWIS, M. M. *Infant speech.* London, Kegan Paul, 1936.

LURIA, A. R., & YUDOVICH, F. I. *Speech and the development of mental processes in the child.* London: Staples Press, 1959.

McCARTHY, D. *The language development of the preschool child.* Minneapolis: University of Minnesota Press, 1930.

MORLEY, M. E., & COURT, D. *The development and disorders of speech in childhood.* Edinburgh: Livingstone, 1957.

PIAGET, J. *The child's conception of number.* London: Routledge & Kegan Paul, 1952.

PIAGET, J. *The child's construction of reality.* London: Routledge & Kegan Paul, 1955.

PIAGET, J. *Proc. XVIth Int. Congr. Psychol.*, Bonn, 1960. [1962]

PIAGET, J., & INHELDER, B. *The child's conception of space.* London: Routledge & Kegan Paul, 1956.

SHERIDAN, M. D. *The child's hearing for speech.* London: Methuen, 1948.

TERMAN, L. M. & MERRILL, M. A. *Measuring intelligence.* London: Harrap, 1937.

TEMPLIN, M. C. *Certain language skills in children.* Minneapolis: University of Minnesota Press, 1957.

VALENTINE, C. W. *The psychology of early childhood.* London: Methuen, 1942.

VERNON, M. D. The relation of cognition and fantasy in children. *Brit. J. Psychol.*, 1940, *30*, 273.

VERNON, M. D. *A further study of visual perception*, Chap. II. Cambridge: Cambridge University Press, 1952.

WALTERS, A. A genetic study of geometrical-optical illusions. *Genet. Psychol. Monogr.*, 1942, *25*, 101.

WITKIN, H. A. *Proc. XVIth Int. Congr. Psychol.*, Bonn, 1960. [1962]

24

Piaget and the Role of Guided Experiences in Human Development

Ralph H. Ojemann and Karen Pritchett

Can the course of development be modified by planned experiences? How fixed are the ages and stages in concept development, for example? Ojemann and Pritchett use guided learning experiences to test a portion of Piaget's theory on cognition. This type of experiment should be contrasted with Elkind (7), Osler and Weiss (41), and Wohlwill and Lowe (25) both in design and theoretical assumptions. Studies such as this one, although difficult to design rigorously, are of great significance not only because they increase our understanding of development but also because they have implications for education.

The purpose of this paper is to consider the nature of "guided" or "planned learning experiences" and to attempt some clarification of the problem of their role in human development. In years past considerable confusion has appeared relative to this question, and the recent revival of interest in Piaget's work on the development of concepts appears to have increased the confusion.

Before proceeding, it is necessary to define the terms *guided experiences* and *planned learning experiences*. In this paper the terms are used synonymously to designate those attempts to assist a child or adult in learning (such as in learning the meaning of a concept or in cultivating an appreciation) by developing, from an analysis of the nature of the learning task and the nature of the learner, a planned sequence of experiences for mastering the task and by motivating the person to participate in this experience. The planned sequence starts with the present development of the learner and leads step by step to the objective.

The term *unplanned experiences*, as used in this paper, refers to the experiences the environment happens to provide. The kind and sequence of experiences are determined by criteria other

---

From *Perceptual and Motor Skills*, 1963, *17*: 927–940.

than the requirements suggested by an analysis of the specified learning task. We do not know in detail the nature of the experiences and background of *Ss* examined by Piaget (1951) in his study of the development of such a concept as specific gravity, for example; but it is highly probable that there was little in the experiences of those children that was based on an analysis of what the learning of the concept "specific gravity" involves.

In contrast to the unplanned experiences, we may consider a planned sequence, as defined above, for learning a given task. For example, an analysis of what is involved in the development of the concept "table" suggests that *S* has to abstract "flat top" supported by something akin to "legs" from other characteristics of objects, such as shape, color, material of which they are made, and to learn to apply the symbol "table" to the combination of "flat top with legs." A planned learning program would take cognizance of this analysis and would consider how the child may be assisted in learning to make this abstraction. The beginnings of such a program might be developed as follows. Suppose the child has been using the words "flat top" and "legs" in such a way as to indicate that these words have considerable meaning for him. We could then consider briefly with the child how it helps us to know what tables are; and, when he shows by his questions and actions that he is ready to proceed, we can provide some such demonstration as the following: "Here is a thing that has a flat top and some legs which hold up the top. (We may point to the flat top and run our hand over it and point to the legs.) Such a thing we call a table. Here is another thing. It also has a flat top and legs. (The pointing may be repeated by both leader and child.) Here are some others. You see they are of different colors and shapes. They are made from different kinds of material. Here are some pictures of more things we call tables.

"Here are some things that do not have legs but the tops are held up by several drawers. They are not called 'tables.' They are called 'desks.'

"Now let's see if we can pick out in these pictures things that are tables and those things that are not."

The demonstration-discussions could proceed with a variety of objects and could include such exercises as sorting tables from other objects. *Ss* could be asked to draw some objects that are tables and some that are not. There would be a variety of experiences from which highly individualized sequences could be built. If a high degree of replicability is desired, the demonstration-discussion could be set up as a machine programmed sequence.

It appears to the writers that there is a growing recognition of the importance of such planned or guided experiences. The increased emphasis on the development of programs for learning a wide variety of complex tasks provides some evidence for this. However, there are numerous investigations in child development which proceed as if planned experiences were not important. The result is considerable confusion.

PIAGET AND EPIGENESIS OF THOUGHT STRUCTURE

For example, an upsurge of interest in the work of Piaget has recently appeared. Hunt (1961) in an extended analysis of Piaget's work points out that Piaget's third theme is

> that the epigenesis of behavioral and thought structures derive from the invariant functions of accommodation and assimilation that operate in the child's continuous interaction with his environment. . . .
>
> Piaget's conception of organism-environment interaction through assimilation and accommodation is neither hereditarian nor is it environmentalistic; it is both. The role of genetic influence is never denied, but it falls far short of being the whole story. In any given situation, the first response of the child is one of those behavioral structures (schemata) already present from past assimilation. What variations in the environment do is to force the child to cope with this variation, and, in the coping, to modify the structures. This latter is accommodation, and the modifications are then assimilated through repetition in practice play (1961, p. 258).

It is well-known that experiences can be of great variety. If thought structures are affected by experiences and if the latter can be of great variety, it seems that any account of the interaction of organism and environment would have to specify the kinds of experiences the children whose growth of concepts was being investigated had had.

Piaget proceeds, however, to delineate stages (with corresponding ages) in the development of concepts of length (1960), conservation of mass (1952), number (1952), physical causality (1951), and others without considering the child's experiences. Wohlwill (1962) comments: "Piaget, as is well known, tends to ignore the effects of antecedent conditions and environmental variables in development, relegating them to a place definitely subsidiary in importance to the unfolding of internal structures" (1962, p. 104).

Not only can experiences be of great variety, they may also act to produce development in opposite directions. For example, in the studies of the writers and their associates (1955) of concepts of social causality, there is evidence that under the ordinary conditions prevailing in the midwestern culture of the United States a tendency exists for the child to develop an arbitrary judgmental orientation toward the social environment. This development, however, can be significantly changed in a more causal direction by the use of planned learning experiences, as indicated by the comparison of experimental and control groups (Ojemann, *et al.,* 1955). If Piaget were investigating the course of development of social causality concepts in the control groups, he would obtain one result. If he used the experimental groups, he would obtain another.

There appear to be at least two aspects represented in Piaget's analysis of the development of concepts. One is related to the ages at which specified stages appear. The other relates to the types of stages that appear and the order in which they appear. For example, in summarizing the data relative to the development of the concept of specific gravity as applied to the floating of boats, Piaget writes:

> During the first stage, which ends at about 5 years, floating is explained by animistic and moral reasons. During a second stage, extending on the average from 5 to 6, the child thinks that boats float because they are heavy. During a third stage (average age 6–8), on the contrary, the child says that they float because they are light. But this explanation is still fraught with reasons of a dynamic order. Finally, at about 9 years, the child begins to understand the true relation between the weight of the boat and that of the liquid element.
>
> Moreover, the earlier the explanations are, the less simple we shall find them. There is an overdetermination of factors which is very interesting to observe. In addition, it should be noted at once that, in spite of appearances, these experiments on floating are subject to a schema of evolution completely analogous to that which we found in our investigation of explanations relative to the suspension of clouds and the heavenly bodies, and to the fall of heavy objects to the ground (1951, pp. 135–136).

In this summary, both the ages at which various stages appear and the order in which they appear are given. Furthermore, there is a suggestion that the order of the development of the concept as applied to the floating of boats also appears in the child's explanation of the "suspension of clouds and heavenly bodies and to the fall of heavy objects to the ground" (1951, p. 136).

INVESTIGATION OF DEVELOPMENT OF "SPECIFIC GRAVITY" CONCEPT

To provide some material for examining further both of the above aspects of the concept of specific gravity, an experiment was conducted using groups of kindergarten and first grade children. The purpose was to test the hypothesis that planned learning experiences can alter the course of development of the concept. The study involved the development of a program of guided experiences to aid the child in developing the concept of specific gravity, administering this program, and testing the results.

The development of a program of guided experiences logically requires a theory of planned learning programs. There is needed a theoretical conception of what goes on in the central mental processes when a planned sequence of experiences is applied and of how the use of a planned sequence of stimuli can bring about the development of a concept.

The theoretical considerations employed in this study are as follows. When a child becomes aware either through external or internal stimuli of the question, *what makes things float?*, a variety of excitations may characterize the central mental processes, depending upon *S's* past experiences with such words as "float" and with objects and experiences related to this area. A functional concept of "why objects float or sink" requires the arousal of the excitations represented by "weight of object," "weight of equivalent body of water," and by the process of relating one to the other. The question *"why does this thing float (or sink)"* should call forth these excitations.

Thus, the development of this concept requires that the aspects represented by "weight of object" and "weight of equivalent volume of water" be abstracted from other aspects of the object such as shape, color, form, and the like. The question of "floating and sinking" must ultimately produce excitations which have a minimum of reference to color, form, and similar aspects and a maximum of reference to "weight of object" and "weight of equivalent volume of water."

To guide the child in making this abstraction, a series of stimuli may be used; namely, words read (verbal stimuli), words spoken (auditory stimuli), demonstrations observed (visual and auditory stimuli), and activities performed (kinesthetic stimuli). These are so arranged that the excitations needed to form the combination represented by the meaning of the new concept are aroused simultaneously or in close succession. The process of arousing the excitations simultaneously must be repeated to give time for the combination of the essential aspects of the concept to form and to have a new symbol attached to this concept.

It was recognized at the outset in the development of a planned program for learning the meaning of specific gravity that use of a number of verbal stimuli would be convenient. To make certain that the words used would arouse the appropriate excitations, the child was tested to determine which key words had meaning for him and which did not.

It was also recognized that internal as well as external stimuli may be operative at any given time. The child is an organism that is aroused to action not only by external stimuli—there are internal stimuli to be considered, also. Thus, the excitations aroused by the external stimuli will vary somewhat from child to child depending upon what internal stimuli are operating at the time. Interference with learning may occur if there are strong internal conflicts (feelings of insecurity, inadequacy, or the like) or if the child feels he is being asked to devote his time to something that he does not consider worthwhile or that represents a threat to his security (questions too difficult) or self-respect (worthwhileness of task), for example. The excitations aroused by the various external stimuli may then produce emotional reactions which interfere with the arousal of the two clusters of meanings—weight of object and weight of equivalent volume of water—and with the formation of a combination to which the words "float" and "sink" can be attached.

Thus, in the early steps of the learning program, there was not only a testing to determine whether the child knew the meaning of "weight," "amount of water," "heavier," and "lighter," but also a probing as to how the child felt about the importance of learning why things float. Did he feel that it might help him to become a more capable person if he knew about the floating and sinking of objects?

Since it seemed desirable to include the series of tests used by Piaget as one measure of the effectiveness of the guided experiences, an attempt was made throughout the learning program, as well as in the pretest, to use objects different from those used in the Piaget series. Thus, such objects as marbles, clay, and nails listed by Piaget were not used in the pretest, the learning program, or the first posttest.

DESCRIPTION OF THE PLANNED LEARNING PROGRAM

The learning program began with a short discussion of the question, *why should we spend our time trying to figure out why things float?* This short discussion brought out three suggestions: (a) we want to know how things work; (b) it helps us to know what will happen when things are put into water (e.g., ice cube into a glass full of water); and (c) it might help us to find a way to learn about other things (problem-solving approach).

Then a plastic bowl approximately 5 in. in diameter was held up, put on the water, and the question asked, *why do you think it would float?* The answers to this question were tabulated (1) because it is light (or heavy); (2) because it is made of plastic (or something that floats); (3) because it is soft (or hard); (4) because it is big (or little).

Each answer was taken in turn and examined. Thus the first response, "because it is light (or heavy)," was examined by using a simple laboratory balance scale to develop the idea that "light" and "heavy" require comparison of at least two things. When the idea of which of two objects is "lighter" and which "heavier" had been discussed, the five pairs of objects were placed on the balance as the entire group observed to determine which was "lighter." The following pairs of objects were used:

Object	Description or approximate size
(a) rectangular piece of iron	3½ in. × 1½ in.
metal jar lid	3½ in. diameter
(b) plastic lid	2 in. diameter
dime	same weight as lid
(c) wooden block	6 in. × 2 in. × 2 in. and heavier than iron
iron piece	same as in pair (a)
(d) die (one of a pair of dice)	1-in. cube
small glass bottle	1½ in. × ½ in. diameter, lighter than die
(e) die	same as in pair (d)
flat plastic toy	same weight as die

Both members of each pair were placed in the water with the group observing the result. The results were tabulated on the board and the question asked, *was our first answer, "because it is lighter (or heavier)," correct?* The group could easily observe that this reason did not produce consistent results and another answer would be needed.

In like manner the other three responses were examined and the session ended with the observation that "we have to find new answers—the old ones didn't work."

At the next session examples were given to evoke the generalization that two things cannot be in the same space at the same time. It was demonstrated that when objects are put into water "the water is pushed away" to make room for the object and that larger objects "push away" more water than smaller objects when each object is put "all the way under the water." Equipment was provided for each of the children to compare the amount of water pushed away by several large and small objects.

Then it was demonstrated that when an object is put in a small vessel of water an observer can see the water "pushed away" rise in the vessel but when it is put in a large vessel of water, it is more difficult to see the water "pushed away." Close observation shows that the water rises.

At the beginning of the third session, it was suggested that "we might see if the weight of the water pushed away has anything to do with the problem." A 1500-ml. glass beaker with a spout on the side about two-thirds of the distance from the bottom was provided. The beaker was filled with water to the level of the spout and the following six objects immersed:

Object	Description or approximate size
(a) a tube of shampoo	4 in. (plastic tube)
(b) a piece of iron	3 in. × 1½ in.
(c) a metal spoon	teaspoon
(d) a metal jar lid	3½ in. diameter
(e) a can of tomato paste	6 oz.—3½ in. × 2 in. diameter
(f) a closed-end aluminum salt shaker same size as can	3½ in. × 2 in. diameter

The water displaced was caught in a light plastic container and placed on one of the pans of the balance with the object in a similar container on the other pan. The class assisted in marking a chart to show whether the object floated or sank and whether the water displaced weighed more or less. The results were observed and the findings verbalized.

Finally, a series of eight exercises in demonstration form were presented to test whether the group could estimate (a) which one of two different sized objects will "push away" more water; (b) which one of a pair of equal-sized objects will displace more water; (c) which will weigh more, the object or water it displaced (given the information that the object would float); (d) which of two objects will float (given their relative weight); (e) whether a hollow keg would float after being filled with material (given a comparison of its weight and that of a similar volume of water).

It may be well to note that the 1-hr. sessions did not seem unduly long for these age groups when interesting materials and varied experiences were provided.

Procedure

The three session learning program was administered to kindergarten and first grade groups. At the kindergarten level a few *S*s were chosen at random for a preliminary individual try-out of the learning program. A random half of the remainder of the class was placed in the experimental group and a random half in the control group.

The pretests and posttests were administered individually with only *E* and the child in the room. The items used in the O-P tests (for Ojemann-Pritchett, to designate a series different from the Piaget test) and the questions asked for each item were as follows:

Objects	Questions
(a) clothespin, cork, candle, eraser, plastic ball, plastic boat, pencil, and rod	"I am going to put this . . . on the water. Will it float, or will it sink? Why?"
(b) large block and water in glasses of same size	"If you put this block on one side of the balance scale and the water on the other side of the scale, which side will go down? Why do you think so?"
(c) red rubber ball	"Would this ball float in this glass of water? In the river? In the ocean? Why would (or wouldn't) it float in each of these places?"
(d) iron washer red rubber ball	"If I put this into the water, will anything happen to the water? What? Why did the water go up in the glass? Why did the water go up more when we put the ball in than when we put the washer into the water?"

In the Piaget series, the following items and questions were used:

Objects	Questions
(a) wooden boat, matchstick, marble, pebble, nail	"I am going to put this . . . on the water. Will it stay on top, or will it go to the bottom? Why?"
(b) piece of wood and equal volume of water	"Which weighs more, this piece of wood, or this water? Why?"
(c) pebble, piece of wood	"I am going to put this pebble/wood in the water. What will happen to the water? Will the water stay in the same place or not? Why? Why does the water rise more for the wood than for the pebble?"
(d) wooden boat	"Would this boat float in this basin of water? Why? In the glass? In the lake? Why would it float in a . . . and not in a . . . ?"
(e) clay	"Will this lump of clay float, or will it sink? Can you make a boat from it so that it will float? (If *S* said "no," or was unable to make the clay float, *E* demonstrated.) "Why will the boat float if it has thin sides, and sink if the sides are thick?"

After a question had been asked, the child was allowed as much time as he seemed to need for his answer. The total time for the administration of the pretest series (O-P test) varied from 7 to 15 min. with an average time of about 12 min. The time for the administration of the posttests (O-P and Piaget series) varied from approximately 13 to 25 min. and averaged about 18 min. The responses were recorded verbatim on the interview schedule developed for that purpose.

The pretest was administered three or four days before the learning program began and the posttests were administered one or two days following the close of the program.

The responses from 10 kindergarten pupils, used in the preliminary testing of the learning program and not included in the experimental and control groups, were scored by two observers using the four stages suggested by Piaget (1951). The definitions of each stage had been worked out from the descriptions contained in Piaget's writings. An agreement of approximately 80% was obtained.

Results: Kindergarten

The results are given as individual scores in Table 1 and in summary form in Table 2. In the analysis presented in Table 2, *t* for small samples was used. An inspection of the data in the table indicates that the two groups did not differ significantly in IQ or in pretest scores. The experimental group made a significant gain over the control group on the posttest series and this gain was also evidenced in the Piaget series. It will be recalled that the materials used in the training program and the O-P posttest series were different from those of the Piaget series. Thus, there was some evidence of generalization.

Since the distribution of scores, especially on the posttest series, was skewed and since this may affect the *t* test, a further analysis was made using the distribution of individual changes from pretest to posttest and testing the significance of the mean gains for experimental and control groups separately. The mean change of the experimental *S*s on the O-P series was 1.3, and *t* was 3.06 which is significant at the .02 level. The mean change on the Piaget series was 1.2 and *t* was 3.39 which is significant at the .01 level. The changes in the control group on the O-P test series and on the Piaget series were

Table 1: *Individual data: Experimental and control groups, kindergarten*

S No.	CA	IQ	O-P Pretest	O-P Posttest	Piaget Test	Individual Change
Experimental Group						
1	5-10	127	2	4	4	2
2	5-4	103	1	1	1	0
3	6-2	119	3	3	3	0
4	6-0	122	1	4	4	3
5	5-7	129	3	3	3	0
6	6-0	116	3	4	4	1
7	6-0	116	2	4	4	2
8	5-7	97	3	4	4	1
9	5-4	117	1	4	3	3 or 2
Control Group						
1	5-7	110	1	1	3	0 or 2
2	5-3	113	1	1	1	0
3	5-9	119	3	3	3	0
4	5-7	112	3	3	3	0
5	5-5	112	1	1	1	0
6	6-3	101	1	1	1	0
7	6-4	100	3	3	3	0
8	5-8	108	3	3	3	0
9	5-8	108	3	3	3	0

Table 2: *Differences between experimental and control group means, kindergarten*

Measure	Exp. Group		Control Group		Difference	t	p*
	M ($N = 9$)	SD	M ($N = 9$)	SD			
IQ	116.2	9.8	109.2	5.6	7.0	1.76	< .10
Pretest	2.1	0.88	2.1	1.0	0.0	0.00	
Posttest	3.4	0.96	2.1	1.0	1.3	2.67	< .02
Piaget Test	3.3	0.94	2.2	1.0	1.1	2.28	< .05

* *Significance data for all analyses are for two-ended critical regions.*

clearly insignificant. Thus, both types of analysis confirm the finding that the kindergarten group made significant growth on both series.

An examination of the individual O-P posttest scores in Table 1 indicates that 6 of the 9 *S*s exceeded the age achievements suggested by Piaget. Seven *S*s exceeded the suggested age achievement on the Piaget posttest. Thus, the ages at which the various stages appear as suggested by Piaget require reinterpretation for American children under the present circumstances.

The question as to whether the order of development is an inherent order, that is, whether the child goes from stage 1 to stage 2, then to stage 3, and finally to stage 4 as outlined by Piaget is more difficult to test critically. One *S* in each posttest series showed a gain of three stages. Two *S*s on the O-P posttest series and three on the Piaget series gained two stages. Did the child who showed a gain of three stages go through each of the three stages during the three-day program, or did the *S* under the influence of the planned experiences modify his thought structures (Piaget's accommodation) from stage 1 to stage 4 without going through the intervening stages? Similarly, did the two (or three, depending on the posttest series considered) *S*s who showed a change of two stages make the modification without going through the intervening stages? The investigators were aware of these questions when the learning pro-

Table 3: *Individual data: Experimental and control groups, first grade*

S No.	CA	IQ	O-P Pretest	O-P Posttest	Piaget Test	Indiv. Change
Experimental Group						
1	6-11	116	1	3	2	2 or 1
2	6-10	116	3	3	3	0
3	6-10	116	3	3	2	0 or –1
4	7-5	113	3	4	4	1
5	6-10	119	3	3	3	0
6	6-7	118	3	4	4	1
7	7-0	122	2	3	3	1
8	6-7	112	3	4	4	1
9	6-8	119	3	3	3	0
10	7-0	110	3	3	3	0
11	6-11	116	3	4	4	1
Control Group						
1	7-5	125	1	3	1	2 or 0
2	6-10	116	3	3	3	0
3	6-7	99	3	1	1	–2
4	6-8	116	3	3	3	0
5	6-7	131	3	3	3	0
6	7-4	96	3	3	3	0
7	6-6	124	3	3	3	0
8	6-5	131	1	1	1	0
9	7-4	107	3	3	3	0
10	6-9	129	3	3	3	0

Table 4: *Differences between experimental and control group means, first grade*

Measure	Exp. Group		Control Group		Difference	*t*	*p*
	M (*N* = 11)	*SD*	*M* (*N* = 10)	*SD*			
IQ	116.1	3.4	117.4	12.3	–1.3	–0.32	> .50
Pretest	2.73	0.6	2.6	0.8	0.13	0.40	> .50
Posttest	3.36	0.5	2.6	0.8	0.76	2.5	< .05
Piaget test	3.18	0.7	2.4	0.92	0.78	2.07	< .06

gram was being administered, but it was not possible in the discussions to follow each child's thinking individually. A more critical test of this question is being sought.

Results: First Grade

Data relative to the experimental and control groups at the first grade level are presented in Tables 3 and 4. Table 3 provides the individual data and Table 4 the summary.

In the first grade group there were no significant differences between the experimental and control groups in IQ and pretest scores. The experimental group showed a gain significant at the .05 level of confidence on the O-P test series and a gain on the Piaget series which approximated significance (t = 2.07 which is just short of the 2.09 needed for significance at the .05 level).

Using the difference scores, the mean was .64 on the O-P series and .45 for the Piaget series. The corresponding ts are 3.60 and 2.16 which are significant at the .01 and .05 levels, respectively. The control group made practically no change. Thus, in this group also there was evidence of significant growth.

Since the oldest *S* in both the experimental and control groups at the first grade was 7 yr. 5 mo., all *S*s scoring four exceeded the age achievement suggested by Piaget. In both the O-P and Piaget posttest series, four (36%) of the experimental *S*s exceeded the level suggested by Piaget.

Since in the first grade experimental group 9 (81%) *S*s scored three in the pretest series, it was not possible for the vast majority of the *S*s to gain more than one step. Thus, this group did not supply data for testing the order of the stages.

DISCUSSION

The above results suggest that in explaining human development one cannot ignore the nature of the child's experiences. Nevertheless, as Gordon (1962) points out, interpretations of developmental data are often made in the child development literature without reference to the child's experiences. It appears from the data reported here that the child's "thought structures" are significantly affected by his experiences. The processes of assimilation and accommodation postulated by Piaget can operate in conjunction with a wide variety of experiences, and it is important for the understanding of that development to specify the nature of the experiences.

It furthermore appears that some *S*s can gain two or three steps, as defined by Piaget during a learning program consisting of a 1-hr. session on each of three successive days. Such results raise the question as to whether the order of development of the stages as outlined by Piaget is a necessary or inevitable order. It was not possible in this investigation to give a decisive answer to this question, but the results suggest that it may not be necessary for the child to go through each step. One of the *S*s went from stage 1 in the pretest to stage 4 in the posttest and three *S*s made gains of two stages.

Two other reactions are recorded here. First, before planned experiences can make their contributions, more attention has to be given to the development of a theory of how they operate. Analysis of the problem suggests that one of the factors contributing to the retarded recognition of the role of guided learning programs may be that investigators working with such programs have not given much attention to the development of basic theoretical conceptions upon which they are based. The development of such conceptions may lead to a deeper understanding of the findings as to relative effectiveness of different methods and similar problems.

Finally, in interpreting the results of the effects of guided experiences, the nature of guided experiences has to be recognized. Guided experiences are planned in the mind of man. In the planning, knowledge of the nature of the organism and the nature of the learning task are utilized. As knowledge of the nature of the organism and of the learning task grows and as more imagination and intelligence are applied to the task, it is highly probable that more effective experiences can be devised. Thus, at any given time the learning potential of a given organism or the course of development has to be interpreted in terms of the quality of the experiences provided. It is only after a variety of experiences have been tested that we can begin to say that a given set of stages and ages is basic or inevitable. Whether children develop according to certain stages may depend on their experiences. Also, questions relative to generalization can be answered only in terms of the experiences that have been "up to now" provided. What generalizations the child is capable of and what bases he can use can be answered only by investigating a wide variety of experiences.

REFERENCES

GORDON, I. J. *Human development*. New York: Harper & Row, 1962.

HUNT, J. McV. *Intelligence and experience*. New York: Ronald, 1961.

OJEMANN, R. H., LEVITT, E. E., LYLE, W. H., JR., & WHITESIDE, M. F. The effects of a "causal" teacher-training program and certain curricular changes on grade school children. *J. exp. Educ.*, 1955, *24*, 95-114.

PIAGET, J. *The child's conception of physical causality*. New York: Humanities Press, 1951.

PIAGET, J. *The child's conception of number*. London: Routledge & Kegan Paul, 1952.

PIAGET, J., INHELDER, B., & SZEMINSKA, A. *The child's conception of geometry*. New York: Basic Books, 1960

WOHLWILL, J. F. From perception to interference: a dimension of cognitive development. *Monogr. Soc. Res. Child Develpm.*, 1962, *27*, No. 2 (Whole No. 83).

25

Experimental Analysis of the Development of the Conservation of Number

Joachim F. Wohlwill and Roland C. Lowe

Piaget's work serves as a source for still another type of evaluation of children's conceptual development. This investigation by Wohlwill and Lowe is an excellent example of the utilization of rigorous experimental design to test various theoretical explanations of learning number conservation. This area of research has grown increasingly important as it has become linked with the teaching of mathematics and the sciences. Along with this article, the reader might wish to examine Osler and Weiss (41) and Ojemann and Pritchett (24).

In Piaget's theory of intellectual development (1950), a central role is assigned to the child's conceptualization of the principle of "conservation," i.e., his realization of the principle that a particular dimension of an object may remain invariant under changes in other, irrelevant aspects of the situation. For instance, children who lack conservation will assert that the relative weight of two objects has changed when the shape of one of them is altered or that numerical equality between two collections of objects no longer holds following a change in the length over which they extend. This phenomenon, which has been demonstrated for a variety of other dimensions, including those of volume, area and length, represents, according to Piaget, a manifestation of the immature level of functioning of the child's mental processes and of their failure to conform to the operational structures of logical thought.

Although Piaget has described some of the precursors of this notion of conservation in children who have not yet attained this level, little is known thus far about the specific ways in which the transition from lack of conservation to the presence of conservation takes place. It is apparent, however, that an adequate explanation of this problem ultimately requires a clearer understanding of the psychological processes at work in this transition phase.

One approach to this goal is to expose children presumed to be slightly below the age of onset of conservation to selected, systematically manipulated learning experiences, designed to call into play different factors believed to be important in the development of conservation. Any differential changes in the children's tendency to give conservation responses should then reflect the role played by the particular factors manipulated. At the same time, a more detailed examination of the interrelationship among different tasks involving conservation, and closely related concepts should likewise extend our understanding of the nature of this problem.

The domain of number lends itself particularly well to the investigation of the development of conservation, for several reasons. First of all, recent empirical work has given strong support to the notion that the attainment of the level of conservation marks a clearly defined stage in the formation of the number concept. Secondly, in this domain the problem of conservation can be readily related to development in other aspects of the number concept (e.g., counting, arithmetical skills, etc.), rather than constituting the somewhat isolated, *sui generis* problem which conservation appears to represent for such dimensions as weight or volume. Thirdly, and most important, the number dimension occupies a unique position in regard to the question of conservation, insofar as the number of elements contained in a particular collection is exactly identifiable by the corresponding integer; by the same token, the *fact* of conservation—i.e., that the number of a collection remains invariant under changes in the spatial arrangement of its elements—is readily verifiable, through the operation of counting. This feature creates an opportunity for assessing the role of symbolic, mediational processes, as well as of reinforcement, in the development of conservation.

This very uniqueness of the number dimension represents of course a potential limitation, as regards the applicability of the results to the problem of conservation in general. It is of considerable interest, therefore, that a rather similar investigation of the acquisition of the conservation of weight has simultaneously been carried out by Smedslund (1961); its results will thus provide us with a valuable basis for comparison, as we will note in the discussion section.

THREE ALTERNATIVE THEORETICAL VIEWS OF NUMBER CONSERVATION

If one looks closely at the problem confronting

From *Child Development*, 1962, *33*, 153–167. Reprinted with permission of the Society for Research in Child Development.

Table 1: *Design of the study*

Order	First Day	Order	Second Day
1.	Diagnostic Questions	1.	Training Series (trials 10 to 18)
2.	Verbal Conservation Pretest	2.	Nonverbal Conservation Posttest
3.	Pretraining in Number Matching	3.	Verbal Conservation Posttest
4.	Nonverbal Conservation Pretest		
5.	Training Series (trials 1 to 9)		

the child in the conservation situation, several different interpretations of the acquisition of this principle suggest themselves. We may label these alternatives the reinforcement hypothesis, the differentiation hypothesis, and the inference hypothesis.

The *reinforcement hypothesis* would propose that, as a child obtains increasing experience in counting numerical collections of different types and in different arrangements, he gradually learns that alterations in the perceptual dimensions of a set do not change its number, i.e., that the same number is obtained from counting the set after as before such a change. Accordingly, systematic reinforced practice in counting rows of elements prior and subsequent to changes in the length of the rows should promote conservation.

The *differentiation* hypothesis would interpret lack of conservation in the young child as a response to an irrelevant but highly visible cue (length) which typically shows substantial correlation with that of number. The child thus has to learn to differentiate the dimension of number from this irrelevant cue. Repeated experience designed to neutralize the cue of length, and thus to weaken the association between it and the dimension of number in the child's thinking, should be expected, then, to facilitate conservation responses.

The *inference* hypothesis, finally, is based in part on Piaget's own analysis of the role of learning in the development of logical operations (1959). Piaget maintains that experiential factors can only become effective, in this realm of development, to the extent that it builds on the child's previously developed structures of thought, as through the activation of a reasoning process prior to, but logically related to the one to be developed. In the case of conservation, one possible implication might be that by dint of cumulative exposure to the effects of adding an element to a collection, or subtracting one from it, the child may be led to *infer* conservation as the result of a change involving neither addition nor subtraction. This implication is supported, incidentally, by the explanations frequently voiced by children who admit conservation, e.g., "it's still the same, because you haven't taken any away."

Prior work by the senior author also bears on this last alternative. First, in the course of a sequential analysis of the development of the number concept (Wohlwill, 1960) it was found that success on tasks involving simple addition and subtraction not only regularly preceded success on a task embodying the principle of conservation, but appeared, in a certain number of subjects, to lead to the emergence of conservation responses. In a subsequent pilot study (Wohlwill, 1959), it was found, furthermore, that subjects given a limited set of trials involving addition and subtraction subsequently made more conservation responses than subjects given equivalent training on conservation, though the difference did not reach significance.

The results of this pilot study suggested the possibility of a more extended investigation of the development of the conservation of number, by bringing to bear each of the above-mentioned theoretical interpretations in the context of a small-scale learning experiment and determining the effectiveness of the various conditions of learning in bringing about conservation, both in a limited and a generalized sense. This is the main aim of the study to be reported, a subsidiary purpose being to provide information regarding the cross-situational generality of number conservation and its relationship to other types of number skills.

METHOD

The experiment was conducted in two sessions over two successive days (except for two *S*s, for whom the interval between sessions was two and six days, respectively). The general design called for (a) a predominantly verbal pretest, partly of a diagnostic character, to reveal *S*'s ability to deal with number concepts, and partly dealing specifically with conservation; (b) a "nonverbal" test of conservation given in the form of a series of multiple-choice trials; (c) a training series on tasks presumed to be related to number conservation; and finally (d) a repetition of both the nonverbal and verbal tests of conservation to provide a measure of learning or change with respect to the understanding of this notion. This design is summarized in Table I.

PROCEDURE

Diagnostic Questions

1. *Number Production*. *S* was shown a pile of red poker chips and was told, "Give me six of them."

2. *Number Equivalence*. *E* laid out a row of seven red chips. *S* was told, "Put down just as many of your chips over here (indicating an imaginary row paralleling *E*'s row), as I have here."

3. *Number vs. Length*. *E* laid out a row of six blue chips extending beyond the limits of his own row of seven red chips (*S*'s row being longer than *E*'s). *S* was asked, "Who has more chips, you or I?" If he answered that he had more, but without having counted the chips, he was asked, "How do you know?"

These three questions concerned, respectively, the child's ability to (a) reproduce a particular cardinal number, (b) establish a relationship of numerical equivalence between two collections, and (c) respond to the dimension of number independent of irrelevant perceptual cues (e.g., length).

Verbal Conservation Pretest

4. Two rows of seven chips each, one blue and the other red, were placed parallel to each other so that both rows were of the same length, and the chips in one row were directly opposite those in the other. *S* was asked, "Who has more chips, you or I?" This question, hereafter referred to as Q, was repeated for all the items in this part.

5a. *E* then extended the red row in both directions to a length about twice that of the blue row. (Q)

5b. The red row was subdivided into two rows of four and three chips placed parallel to *S*'s blue row. (Q)

5c. The red chips were placed in a vertical pile in front of the blue row. (Q)

5d. The red chips were inserted into an opaque tube. (Q)

Question 4 served chiefly a preparatory function, i.e., to set up the following questions of conservation. Question 5a represented the main criterion of number conservation, while 5b to 5d indicate the generalizability of conservation. Accordingly, questions 5a to 5d were cumulative: if a *S* did not assert equality at any point, the remaining questions were omitted.

6. Questions 4 and 5a were repeated with 12 chips in each row instead of seven.

The suggestive nature of the questions (Q) used above ("Who has more chips, you or I?") requires comment. It should be noted that its initial use (in question 4) is in a situation where perceptual cues mitigate against the child's following the suggestion of inequality implicit in the question: the matched rows of chips afford a strong cue for direct perception of equivalence.[1] Second, the suggestion applied both in the pre- and the posttests and thus may be presumed to have played a constant role on both occasions.

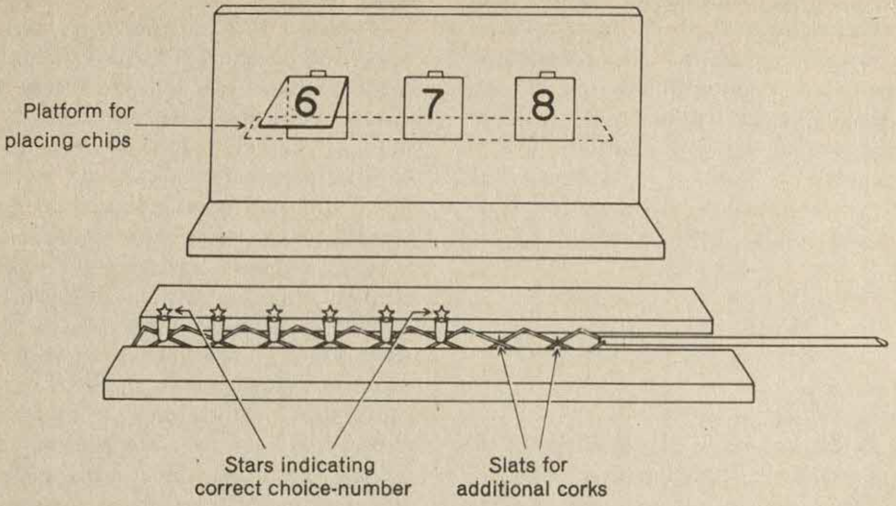
Pretraining in Number Matching

The apparatus used here, shown in Figure I, consisted of an upright panel containing three windows which had the numerals 6, 7, 8 inscribed on them from left to right. The *S*s were told they were going to play a game, in which they would find a chip hidden behind one of the windows, and that the object of the game was to get as many chips as they could. For the pretraining phase, the procedure consisted in presenting singly a series of six 5 by 5 in. cards showing six, seven, or eight colored stars arranged in simple configurations. On each trial a colored chip was hidden behind the corresponding window. *S* was informed that the number of stars on the card would tell him behind which window the chip was hidden and urged to count the stars.

---

[1] All but 14 of the *S*s did in fact resist *E*'s suggestion, usually through some such answer as "we both do" or "you and me." The 14 *S*s who failed to do so were made to count the two rows, whereupon the question was again put to them. If *S* persisted in following the suggestion of inequality, *E* confronted him with the results of his counting and, if necessary, told him outright "So we both have the same, don't we?" This procedure was necessary in order to proceed to the following part, where *S*'s knowledge of the equivalence of the two rows had to be presupposed.

Fig. 1: *Apparatus for nonverbal conservation and training trials, showing device for presenting variable-length rows of sample numbers and display-board for choice numbers.*


Platform for placing chips
6
7
8
Stars indicating correct choice-number
Slats for additional corks

When *S* opened the correct window, he was instructed to remove the chip and place it onto a board at his side. He was told to fill up the board with chips; if he found "a lot" of them, he would on the following day receive a toy. The purpose of this series was to create a set in the child to respond to number, as well as to familiarize *S*s with the specific numbers shown. A correction procedure was used which involved having *S*s correct any mistakes made in counting the stars and guiding *S* to the correct window when, as occasionally happened, a *S* counted correctly but made an incorrect choice.

Nonverbal Conservation Pretest

This series consisted of three two-phase trials. *S*s were presented with a row of colored stars, either six, seven, or eight in number; mounted on a set of corks which rested on a series of connected scissors-like slats. This apparatus, depicted in Figure 1, permitted lengthening or shortening the row while preserving the straight-line arrangement. *E* told *S* that he was to count the stars in order to find the chip behind the correct window. Following *S*'s initial response, he was made to return the chip to *E*, who replaced it behind the same window, and then, depending on the trial, either extended or shortened the row of stars. *S* was allowed to count only on the first phase; he thus had to find the correct window on the second phase on the basis of the knowledge gained in the first and in the face of the perceptual changes in the row of stars.

Training

There were four conditions of training: Reinforced Practice, Addition and Subtraction, Dissociation, and Control. The three experimental conditions were designed to relate, respectively, to the reinforcement, inference, and differentiation hypotheses presented in the introduction. Each training series consisted of 18 trials, broken up into two sets of nine which were administered on successive days. The apparatus used was the same as in the conservation pretest trials.

a. *Reinforced Practice* (RP). The procedure here was the same as for the preceding conservation trials, with this modification: if *S* made an incorrect response on the second phase of the trial, he was told to count the stars, so as to find out which window he should have chosen. *E* then exposed the chip behind that window but did not allow *S* to remove the chip.

b. *Addition and Subtraction* (A&S). These trials were similar to the conservation trials, except that on two-thirds of the trials, following the *S*'s initial response after counting, *E* either added or subtracted a star at the end of the row before changing its length. The remaining third of the series consisted of straight conservation trials which were interspersed with the A&S trials.

c. *Dissociation* (Diss.). Unlike the above, these were single-phase trials, with the length of the row varying from one trial to the next over a range of four times the smallest length. *S* was urged to count the stars and open the corresponding window; if correct, he received the chip. Over the series of trials each number of stars appeared equally often at each of the different settings of length.

d. *Control.* This series of trials consisted likewise of single-phase trials as in the Dissociation condition, but the length of the row remained fixed throughout at its minimum spread.

Posttests

The *nonverbal* conservation posttest, consisting of three conservation trials as in the pretest, followed immediately upon the completion of the training trials. (For *S*s in groups Diss. and Control, *E* prefaced these trials with a remark to the effect that they would again have to find the chips twice in a row, the second time without counting.) Any *S* responding correctly on the last trial of this posttest was asked: "How did you know where to look for the chip that time?"

The *verbal* conservation posttest, consisting of a repetition of questions 4 through 6 as given in the pretest, concluded the experimental session.

At the end of the second session, each child was shown a variety of dimestore toys from which he picked one to take back with him as his "prize." Altogether, each of the sessions lasted about 20 to 25 minutes per child. The children's level of attention and motivation appeared to have remained high throughout these sessions, the "game" aspect of the situation apparently having proved effective in capturing their interest. This was reflected in their universal eagerness to return to it when called for the second session.

SUBJECTS

Subjects for this study were 72 kindergarten children, 35 boys and 37 girls, with a mean CA of 5 years, 10 months. (This age level was selected as one at which most children would still show lack of number conservation, while yet being old enough to be able potentially to profit from the learning experience; in other words, an interaction between learning and developmental level is presumed.) There were 18 *S*s in each condition of training, *S*s being assigned to their group according to a predetermined order. The four subgroups were closely matched as to their mean CA. (The range of the means was one month.)

The children were enrolled in the kindergarten classes of three public schools in Worcester, Massachusetts, located in predominantly lower-middle-class neighborhoods. They thus had been and were being exposed to a variety of activities in the area of number skills, consisting mainly of counting, number-matching, and identifying simple numerals.

RESULTS

The presentation of the results of the experiment is divided into three sections: the verbal pretest, including the diagnostic and verbal conservation questions; the learning of nonverbal conservation; and the transfer of training to the verbal posttest.

Verbal Pretest

Considering first the diagnostic questions, only one of the 72 *S*s failed question 1, while four failed question 2. On question 3, however, only 20 *S*s gave a correct response (i.e., based on counting the chips either before or in justification of their judgments). These results show that the *S*s had adequate facility in counting and dealing with numbers symbolically in simple situations, such as producing a required number of elements and in matching two groups for number. Their success on these two tasks, however, contrasted sharply with their performance on question 3 where the task required the abstraction of number as independent from certain irrelevant perceptual cues. It should be noted that, since this question followed question 2 without a break, some *S*s may have seen *E* take a chip from his row in setting up question 3. Thus, some of the correct responses may have been facilitated by this circumstance. In fact, nine *S*s explicitly based their answers on this cue. (Control over this factor in a subsequent study did indeed result in lowering still further the number of *S*s succeeding on this question,

Table 2: *Performance on conservation before and after training*

Condition of Training	Verbal Conservation*			Nonverbal Conservation†		
	Pretest	Posttest	Net Change	Pretest	Posttest	Net Change
A&S	1	3	+ 2	1.05	1.77	+ .72
RP	2	3	+ 1	1.22	1.50	+ .28
Diss.	4	2	− 2	1.05	1.16	+ .11
Control	2	4	+ 2	1.44	1.96	+ .52

* *Number of Ss giving correct responses on question 5a.*
† *Mean correct responses out of three trials.*

so as to equate it in difficulty with the conservation question, 5a.)

On the verbal conservation items, only nine of the 72 *S*s answered correctly on question 5a. A breakdown of the incorrect responses shows that 41 *S*s responded to the length of the rows, while 22 responded to the density of the elements. This tendency to regard the longer row as more numerous was also found on question 3 of the diagnostic questions.

As for the generality of the *S*s' concept of number conservation, of the nine *S*s who succeeded on question 5a, six extended their conservation to 5b, five to 5c, and four to 5d. On question 6, on the other hand, involving conservation for 12 elements, seven of these *S*s showed conservation, in addition to one who had not responded correctly on question 5a. Thus, when conditions were qualitatively different, generalization was somewhat lower than it was when the new situation differed only in a quantitative way.

Nonverbal Conservation Learning

Table 2 summarizes the performance of each group on the verbal and nonverbal tests of conservation, before and after the learning series.

An analysis of variance revealed no significant differences among training groups with respect to learning of nonverbal conservation ($F = 1.73$; $p > .05$ for 3 and 68 *df*). However, the mean over-all difference scores differed significantly from 0 ($t = 3.95$; $p < .01$), showing that for the total group as a whole conservation did increase from pre- to posttest.

A comparison between the responses of the A&S and RP groups on the conservation trials of their respective training series shows that the former *S*s were correct on 48 per cent of their trials, while the latter were correct on 47 per cent. It will be recalled that only six conservation trials were given in the A&S series, while the RP series consisted wholly of 18 conservation trials for which, in addition, a correction procedure was used. Hence, direct training on conservation was no more effective than the more intermittent practice afforded on the A&S trials.

It was also found that the A&S group had greater success on the A&S trials than on the straight conservation trials: for the former, 59 per cent of the responses were correct, as compared to the 48 per cent for the conservation trials. This finding, which is consistent with the results of previous research, represents of course a prerequisite for the use of the A&S trials as a training experience.

The training trials under the control and dissociation condition, which involved only rote counting, were quite easy for these *S*s: a near perfect performance was the norm.

Transfer of Training to Verbal Posttest

With respect to verbal conservation, there were very few changes in any group. The number of *S*s showing conservation of number on the pretest was nine, while 12 *S*s showed it on the posttest. Two *S*s changed to conservation from the A&S group, two from the Control group, and one from the RP group. Two *S*s, in the Diss. group, who had shown conservation on the pretest, failed to do so on the posttest (cf. Table 2).

It is interesting to note that, whereas on the pretest of verbal conservation only four of the nine *S*s having conservation showed perfect extension of this concept on items 5b through 5d, on the posttest nine of the 12 *S*s showing conservation did show this extension, the remaining three *S*s belonging to the group of five who had not shown conservation on the pretest. This seems to indicate the unstable nature of the *S*s' conservation, as acquired in this situation.

Of the 12 *S*s showing conservation on question 5a of the posttest, 11 again showed conservation for 12 elements (question 6).

DISCUSSION

In this section we will consider some of the more specific implications of the results for the nonverbal conservation learning and for the transfer to the verbal test, leaving until a later section certain more general conclusions suggested by this investigation.

Nonverbal Conservation Learning

As regards the "learning" of conservation within the limited context of the training trials, a significant amount of improvement from pre- to posttest did take place for the group as a whole, but the lack of significant differential effects due to the conditions of training and the fact that the Control group gained more than either the Reinforced Practice or the Dissociation groups clearly prevents us from attributing beneficial effects to any specific learning condition.

The failure of the RP group to outperform the others nevertheless deserves comment. It had actually been anticipated that this group, which received essentially one continuous series of conservation trials, would as a result of this extended practice show the greatest amount of learning from pre- to posttest, although such learning might not necessarily transfer to the verbal posttest. The contrary results bear out the ineffectiveness of continued reinforced practice in bringing about conservation responses, even of a purely empirical sort (i.e., "pick the window where the chip was before"), which the above-mentioned pilot study (Wohlwill, 1959) had already hinted at in a much shorter training series. Whether a still more extended series than that used in the present study might have yielded a greater amount of learning remains an open question, of course.

The greatest amount of improvement from the pre- to the posttest trials, on the other hand, took place in the A&S group, exposed to 12 addition and subtraction trials, set off against six conservation trials; these results are thus at least consistent with the possible role of a process of inference (i.e., conservation as the end-product of changes involving neither addition nor subtraction) to which the previous studies (1959, 1960) had pointed.

Finally, as regards the virtual absence of learning in the Dissociation group, it might be suggested, in retrospect, that the very act of counting the stars interfered with directing the child's attention to the cue of length, which the condition was designed to neutralize. If so, no improvement on the conservation trials, based on explicit disregard of the biasing cue of length, would result.

Transfer to verbal conservation questions.

Perhaps the major finding of the study is that none of the above procedures proved in any way effective in leading to an understanding of the principle of number conservation, such as the verbal posttest demanded. For instance, over the four training groups combined, a total of 10 *S*s shifted from zero or one conservation responses on the nonverbal pretest to three on the posttest, yet these shifts did not bring with them a single change to conservation on the verbal posttest.

In explanation for this failure of the nonverbal conservation learning to transfer to the verbal posttest, one might suggest that the nonverbal learning situation favored the development of an essentially empirical rule, i.e., "the correct number remains the same as before after *E* shortens or lengthens the row," or simply "look for the chip behind the window where the chip was just previously." If this were the case, little if any transfer to the very different situation confronting the child in the verbal conservation questions would be expected. The verbalizations elicited from those *S*s who made a correct response on the last posttest conservation trial lend some support to this argument: many of the *S*s actually gave no meaningful explanation for their choice at all (e.g., "I just knew," or "I thought hard about the stars"), while most of the rest responded in such terms as "It was there before."

Interestingly enough, Smedslund (1961), on the basis of his work on the learning of the conservation of weight, similarly argues for the very limited, nonconceptual nature of such learning. In his study, *S*s were exposed to an extended series of judgments of the relative weight of two masses of plasticine, before and after one of these was deformed in shape; each judgment was reinforced by weighing the two objects on a balance. While *S*s did learn to anticipate correctly the conservation of weight of the deformed object, the author feels that this learning was mainly that of an empirical fact, rather than of a logical principle, as shown both in the kind of explanations offered by the children, and in the lack of transfer of the learning to problems embodying logically equivalent principles (e.g., transitivity relationships). Parenthetically, it is worth noting that in Smedslund's study a training procedure embodying addition and subtraction of matter, in a manner somewhat analogous to that of our A&S condition, yielded nearly as much learning as continued practice on conservation problems.

There remains, however, an alternative interpretation of our results. It is based on a major difference between the nonverbal and verbal tests of conservation, which might itself have accounted for the lack of transfer observed: while the nonverbal test involved a match between a given collection of elements and the corresponding, symbolically indicated number, the verbal test entailed rather the equivalence of the numerosity of two collections of elements. Thus, it is conceivable that the children did in fact learn, in their nonverbal training, that the *absolute* number of elements remained unchanged, without transferring this principle to the *relative* number of elements in two collections, in the verbal test. Implausible as this possibility may seem to a sophisticated adult, it is borne out by the total inefficacy of asking the children to count the two collections after a nonconservation response on the verbal posttest: of 23 *S*s who were asked to do so, 19 persevered in their nonconservation responses when the question was repeated, immediately after ascertaining that there were seven chips in each row. . . .

Finally, the use of nonverbal methods in the investigation of children's thinking deserves brief comment. While the ineffectiveness of the nonverbal training procedures in our study may seem to cast doubt on the fruitfulness of such methods, they have been used to good advantage in several other recent studies; moreover, the pitfalls of the verbal interrogation approach, at least as used by Piaget, have been persuasively analyzed (Berko & Brown, 1960, pp. 536f; Braine, 1959). Perhaps the central point is that it is incumbent on those applying nonverbal methods to determine, by varied and appropriate transfer tests, the breadth and depth of the child's understanding of the principles or concepts in question—a point which appears of special relevance to the application of automatic teaching methods to instruction in this and similar areas.

CONCLUSIONS

Although the predominantly negative outcome of this investigation does not allow us to give any definitive answer to the question posed at the outset, concerning the mechanisms involved in the child's acquisition of the concept of the conservation of number, a few general conclusions regarding this problem may be permissible.

First, the strong tendency of the children in this investigation to respond on the basis of differences in length in making numerical comparisons between two collections, even without the element of perceptual *change* introduced in the conservation situation (cf. question 3), lends some weight to the interpretation of lack of conservation as a failure to differentiate number from irrelevant perceptual cues, pointing to an aspect of the problem which appears to have received insufficient attention in Piaget's theoretical account of conservation.

Second, the consistent tendency across several studies for the A&S conditions to yield the most improvement in nonverbal conservation suggests that a process of inference may be operative in the development of number-conservation, even if this inference may be too limited in scope to lead to a generalized understanding of the principle. In view of the fact that children typically receive considerable experience in simple addition and subtraction in the very time period in which conservation generally appears (i.e., in late kindergarten and early first grade), this factor merits further attention.

Third, our investigation highlights the considerable gap separating the ability to *enumerate* collections by counting from a true understanding of the number concept, as it is reflected in the principle of conservation. In this respect the present results are entirely in agreement with those obtained in previous work on the development of the number concept. Furthermore, even repeated identification of a collection with a particular number symbol, independent of length, appears to be relatively ineffective in bringing about conservation, thus raising the question of the adequacy of a mediation-theory approach to this particular aspect of concept formation.

In a more positive vein, two suggestions for future attacks on this problem might be offered. The first is to construct a set of learning experiences which would not only be more extended but, more important, cover a wider variety of situations (i.e., stimulus materials, configurations, specific numbers involved, etc.). This would be in line with Harlow's (1951) emphasis on *generalized* experience as a prerequisite for the learning of broad concepts and principles in primates as well as in man. It is plausible to suppose, in fact, that it is precisely such generalized experience—in the classroom, at play, and in other everyday activities of children of this age level—which represents the basis for the seemingly spontaneous appearance of conservation in the child.

The second suggestion is to undertake a thorough, intensive analysis of the ontogenesis of conservation in a selected number of children followed longitudinally. Special attention might be paid to the types of explanations given by the child at various stages, as well as to the stability and generalizability of conservation responses once they appear. Inhelder and Noelting, at the University of Geneva, have in fact already launched such a longitudinal project, with preliminary results that appear promising (cf. 1957).

SUMMARY

This study represents an attempt to determine more specifically the nature of the processes at work in the development of the notion of the conservation of number (invariance of number under changes in length or configuration of a collection), as studied by Piaget. The investigation was in the form of a nonverbal matching-from-sample type learning experiment, preceded and followed by verbal questions to measure the child's understanding of the conservation principle. There were four conditions of training, involving respectively the role of reinforced practice on conservation, of dissociation of biasing perceptual cues, and of inferential mechanisms based on the recognition of the effects of addition and subtraction of elements; a control group was also included. Subjects were 72 kindergarten-age children.

The results indicate an over-all increase in nonverbal conservation responses from a pre- to a posttest, within the limited context of the learning task, but they show no significant differences attributable to the conditions of training. Transfer of conservation learning to the verbal posttest was negligible under all conditions, indicating that whatever learning may have taken place was of a rather restricted type, representing perhaps more the formation of an empirical rule than the understanding of a general principle.

These results, together with additional findings pertaining to the relationship of certain number skills to conservation, are discussed in terms of their implications for the problem of the development of conservation.

REFERENCES

BERKO, J., & BROWN, R. Psycholinguistic research methods. In P. H. Mussen (Ed.), *Handbook of research methods in child development.* New York: Wiley, 1960. Pp. 517–557.

BRAINE, M. D. S. The ontogeny of certain logical operations: Piaget's formulation examined by nonverbal methods. Psychol. Monogr., 1959, 73, No. 4 (Whole No. 475).

HARLOW, H. F. Thinking. In H. Helson (Ed.), *Theoretical foundations of psychology.* New York: Van Nostrand, 1951. Pp. 452–505.

INHELDER, B., & NOELTING, G. Le passage d'un stade au suivant dans le développement des fonctions cognitives. *Proc. XVth Int. Congr. Psychol.*, Brussels, 1957. Pp. 435–438.

PIAGET, J. *The psychology of intelligence.* London: Routledge & Kegan Paul, 1950.

PIAGET, J. Apprentissage et connaissance. In J. Piaget (Ed.), *La logique des apprentissages.* Paris: Presses of the University of France, 1959. Pp. 159–188. (Etudes d'épistémologie génétique, X).

SMEDSLUND, J. The acquisition of conservation of substance and weight in children. *J. Scan. Psychol.*, 1961, *2*, 71–87.

WOHLWILL, J. F. Un essai d'apprentissage dans le domaine de la conservation du nombre. In J. Piaget (Ed.), *L'apprentissage des structures logiques.* Paris: Presses of the University of France, 1959. Pp. 125–135. (Etudes d'épistémologie génétique, IX).

WOHLWILL, J. F. A study of the development of the number concept by scalogram analysis. *J. genet. Psychol.*, 1960, *97*, 345–377.

PERSONALITY DEVELOPMENT

26

Dependency and Helpfulness in Young Children

Marjorie Stith and Ruth Connor

This study by Stith and Connor illustrates the use of time-sampling techniques and preset categories in testing the relationship between age and emotional interaction of preschool children. It also illustrates partial application of the procedure advocated by Pease and Hawkes (21). Time sampling is useful in providing data of a normative nature, but it does not allow for the interactional study of the diadic relationship. As the authors indicate, the **how** *and* **why** *of developmental changes are not revealed in this study.*

Little research has been concerned with expressions of helpfulness or cooperation of very young children, or with the relative frequency with which children tend to express dependent and helpful behavior. Many child development specialists have assumed that there must be a satisfying dependency relationship with adults before a child can be secure enough to achieve a good relationship with his peers. Extreme dependency on adults in a nursery school situation, however, has been regarded as evidence that the need of the child for satisfaction in his contacts with adults, particularly his parents, has not been met.

Heathers (1955) has pointed out that both adults and children in our culture "accept" some types of behavior which have been perceived as dependent more readily than they accept other kinds of dependent behavior. Heathers has hypothesized that the need to depend on adults does not necessarily decrease with age; it may be, rather, that inhibitions against expressing such need increase with age.

The present study was designed to explore and compare helpful and dependent behavior in young children in order to discover whether the frequency and proportion of helpful and dependent acts were independent of the age and sex of the subject and whether the frequency and proportion of choice of adults or children as recipients of helpful and dependent contacts were independent of the age and sex of the initiator. Also, the investigation sought to discover whether older children seek to satisfy dependence needs in a different fashion than do younger children.

METHOD

The 30 boys and 35 girls, ranging in age from 38 to 75 months, enrolled in the Child Development nursery school and kindergarten of Florida State University, served as subjects (*Ss*) for the study; direct observation using predetermined categories indicating helpful and dependent behavior and a time-sampling technique provided the data. Twenty two-minute observations, according to an established order, were completed on each child. Behavior was divided into seven categories of helpful (offering information, giving help, giving reassurance, giving permission, giving praise, giving affection, and giving reward) and seven categories of dependent behavior (seeking information, seeking help, seeking recognition, seeking praise, seeking affection, seeking reward, and seeking permission). In each instance the behavior was classified as a contact or a response, according to whether or not the action was initiated by the child being observed. Observer reliability, calculated by determining the percentage of agreement between two observers, reached 98 per cent before the collection of data was begun. The behavior categories were adapted from those described by Moustakas, Sigel, and Schalock (1956).

RESULTS

During the 1300 observations there were 5893 recorded behaviors. No consistent relation between frequency of total behavior and age was observed. . . .

There were no over-all relations between either

From *Child Development*, 1962, *33*: 15–20. Reprinted with permission of the Society for Research in Child Development.

Table 1: *Average frequency and proportion of helpful and dependent behavior of subjects*

	Age in Months				
	38–46 (*N*=11)	*47–55* (*N*=15)	*56–63* (*N*=16)	*64–75* (*N*=23)	χ^2
Helpful Behaviors	36.7	46.0	50.0	53.3	9.25†
Dependent Behaviors	52.4	41.3	41.7	38.9	8.10†
Proportion of Helpful Behavior*	41.8	52.7	54.4	57.8	9.59†

** Proportion of helpful behavior in relation to the combined totals of helpful and dependent behaviors.*
† $p < .05$.

Table 2: *Average frequency and proportion of subjects' contacts with adults and children*

	Age in Months				
	38–46 (*N*=11)	*47–55* (*N*=15)	*56–63* (*N*=16)	*64–75* (*N*=23)	χ^2
Contacts with Adults	16.1	10.7	10.9	11.8	7.79†
Contacts with Children	13.8	12.1	18.4	17.1	7.09†
Proportion of Contacts with Adults*	53.6	46.9	37.0	40.7	9.03†

** Proportion of contacts with adults in relation to the combined totals of helpful and dependent contacts with adults and children.*
† $p < .05$.

frequency or proportion of helpful or dependent behavior and the sex of the *S*s. Nor was there a significant relation between the frequency or proportion of choice of adults or children as recipients of either helpful or dependent behavior and the sex of the initiator. However, several significant relationships were noted between these behaviors and the age of the *S*s.

The median test was used to analyze both the frequencies and proportions of dependent and helpful behavior at the age levels designated. The chi square values were calculated not on the basis of percentages, but on the basis of frequencies of the proportions of helpful behavior that fell above the median proportion (and those that fell at the median and below) relative to the combined totals of helpful and dependent behaviors. Chi square values were also calculated on the basis of frequencies of proportions of (a) helpful contacts with adults and children to the combined totals of helpful and dependent contacts and (b) contacts with adults and children.

Table 1 presents the frequencies and proportions of total behavior responses of the children in relation to age. Helpful behaviors per child increased and dependent behavior responses per child decreased as age increased. Among the older children the proportion of dependent responses decreased while the proportion of helpful responses increased.

Frequency of contacts with adults decreased and frequency of contacts increased with age, but not significantly, as shown in Table 2. *S*s in the youngest age group made more contacts with adults than with children, while *S*s in the oldest groups made more contacts with children than with adults. The proportion of contacts with adults decreased while the corresponding proportion of contacts with children increased with age. This relation was significant at the .05 level.

An analysis of the contacts with adults showed at every age level more dependent than helpful contacts. While the youngest group made almost six times as many dependent and helpful contacts with adults, the oldest group made only twice as many. As age increased, there was a greater frequency of helpful than of dependent contacts with adults. The decrease in dependent contacts was significant beyond the .01 level. Also, the proportion of helpful contacts increased significantly (*see* Table 3).

The contacts made with children also changed

Table 3: *Average frequency and proportion of subjects' helpful and dependent contacts with adults*

	Age in months				
	38–46 ($N=11$)	47–55 ($N=15$)	56–63 ($N=16$)	64–75 ($N=23$)	χ^2
Helpful Contacts with Adults	2.5	3.1	2.6	4.4	5.83†
Dependent Contacts with Adults	13.5	7.6	8.4	7.3	11.96‡
Proportion of Helpful Contacts with Adults*	15.8	28.8	23.4	37.6	7.90†

* *Proportion of helpful contacts with adults in relation to the combined totals of helpful and dependent contacts with adults.*

† $p < .05$.

‡ $p < .01$.

Table 4: *Average frequency and proportion of subjects' helpful and dependent contacts with children*

	Age in months				
	38–46 ($N=11$)	47–55 ($N=15$)	56–63 ($N=16$)	64–75 ($N=23$)	χ^2
Helpful Contacts with Children	5.8	5.2	8.7	7.9	8.36†
Dependent Contacts with Children	8.0	6.9	9.0	9.3	4.42
Proportion of Helpful Contacts with Children*	42.1	43.1	47.8	45.8	1.46

* *Proportion of helpful contacts with children in relation to the combined totals of helpful and dependent contacts with children.*

† $p < .05$.

with age. There was an increase in both helpful and dependent contacts with children in relation to the age of the *S* making the contact, but the increase of helpful contacts only was statistically significant as shown in Table 4. Table 4 shows also that there was no consistent change in the proportion of helpful contacts of *S*s with other children; however, the oldest group had a larger proportion of helpful contacts than did the youngest group.

DISCUSSION

As age increased among the *S*s of the present investigation, total helpful responses and helpful contacts increased significantly, while total dependent responses and dependent contacts decreased. . . .

According to other investigators, Boehm (1947), Heathers (1953), and Sears *et al.* (1953), marked decrease occurred in dependence on adults as age increased. This investigation supports these findings and, in addition, hypothesizes that this change may occur because of the increasing incidence of helpful contacts with adults and increased interaction of children as they grow older. While total contacts with adults decreased, more of the contacts which were made were classified as helpful rather than dependent as age increased. Although a larger proportion of dependent contacts were made with adults than with children in the two younger groups, this situation was reversed in the two older groups who made more dependent contacts with children than with adults.

On the other hand, contacts with children increased with age in both number and proportion. Although both dependent and helpful contacts with children increased, helpful contacts increased more than did dependent contacts. At every age level the *S*s made more helpful contacts with children than with adults, but there was a significant increase in both frequency and proportion of the helpful contacts with children as the age of the initiator increased.

That older children meet dependency needs in a different fashion than do younger children seems to be evident, but why or how this happens was not indicated by the findings of this investigation. The way in which changes occur which enable a child to meet the basic needs within himself and at the same time to take his place in the social group in which he moves has been given little attention by researchers.

SUMMARY

A time-sampling technique and predetermined categories were utilized to determine the frequency and proportions of helpful and dependent behavior exhibited by 65 preschool children during 1300 two-minute observations. While, as age increased, dependent contacts with adults decreased in both frequency and proportion, helpful contacts with both adults and children increased significantly. Although dependent contacts with children also increased with the age of the initiator, this change was not statistically significant. The sex of the initiator was not significantly related to either helpful or dependent contacts with either adults or children.

REFERENCES

BOEHM, L. The development of independence: a comparative study. *Child Develpm.*, 1947, *28,* 85–92.

HEATHERS, G. Emotional dependence and independence in a physical threat situation. *Child Develpm.*, 1953, *24,* 169–179.

HEATHERS, G. Acquiring dependence and independence: a theoretical orientation. *J. genet. Psychol.*, 1955, *87,* 277–291.

MOUSTAKAS, C. E., SIGEL, I. E., & SCHALOCK, H. D. An objective method for the measurement and analysis of child-adult interaction. *Child Develpm.*, 1956, *27,* 109–134.

SEARS, R. R., WHITING, J. W. W., NOWLIS, V., & SEARS, P. S. Some child-rearing antecedents of aggression and dependency in young children. *Genet. Psychol. Monogr.*, 1953, *47,* 135–234.

27

Self-Derogations in Young Retardates and Special Class Placement

Joseph H. Meyerowitz

An important question in the field of special education is the effect of special class placement upon the mentally retarded child. Although no answer is given here, this paper does illustrate the development of an instrument to measure self-esteem in young children. Meyerowitz demonstrates that it is possible to elicit verbal self-estimates from preschool children, thus indicating their early evaluation and awareness of self. This study is a part of a larger project of the University of Illinois Institute for Research on Exceptional Children.

The article reports a segment of a larger investigation into the efficacy of special homogeneous classes for educable (Binet IQ 60 to 85) mentally handicapped (hereafter referred to as EMH) children. It is suggested that the implications of this research are relevant to the special psychology of deviancy and childhood, generally. Advocates of special classes for EMH children contend that, among other advantages, these classes promote in the children the acquisition of a more nearly realistic and healthy self-concept. This contention is based upon the assumption that regular class placement confronts the EMH child with standards so out of reach that he has no substantial basis for self-evaluation. It is further assumed that the EMH child who works in a setting wherein he can operate at a rate that is consistent with his learning characteristics will thereby learn to recognize the reality of his assets and liabilities. Past investigations of academic progress in special classes offer little clear-cut evidence for or against the utility of these classes. However, the bulk of studies tend to support the contention of superior adjustment for the EMH child placed in a special class. Jordan and DeCharmes (1950) suggest that this superior social adjustment may be attributed to the EMH child's fear of failure experienced in a regular grade, due to the pressure for academic

---

From *Child Development*, 1962, *33:* 443–451. Reprinted with permission of the Society for Research in Child Development.

achievement; this pressure is not felt by the child in the special class.

While the argument for special class placement is generally accepted, it overlooks the implications of the placement itself on self-concept. The question must be raised: What impact is there on the youngster's self-perception when he is singled out from a classroom of friends and neighbors and placed in a different room, if not in a different school? Does this action confirm or stimulate any existing feeling of unworthiness?

As will be indicated in the section on procedure, measures were taken in this investigation to control those factors which have complicated the evaluation of efficacy of special classes in previous investigations. These factors are:

1. A tendency to favor for placement in special classes troublesome, disturbed, or lower level children within the IQ range. (School officials did not identify EMH children until they had a history of failure.)
2. Stereotyping by the child's peers of the EMH room, teacher, and/or title of the special class.
3. The complexity of adjustment to intellectual peers who are not simultaneously age-peers.
4. Attempts to simulate the normal curriculum in the EMH class, which encouraged adults, outside of the school system, to judge the EMH child's progress by that of children in the regular curriculum. This resulted in discouragement for the EMH child.

The rationale for the present study is that there are two possibilities of change in self-concept through placement in special class: (a) the child may feel rejected because he is segregated from his peers; or (b) the child may feel now, among intellectual peers, adequate and accepted. If investigation should show that there is no difference between children remaining in their regular grades and children placed in a special class, this would be interpreted to mean that placement in a special class has produced no change in self-concept.

This article reports a study of the effects upon the EMH child's perception of himself produced by placement in a special class. If as Jordan and DeCharmes conclude, i.e., in the regular grades, there is pressure for academic achievement of which especially the EMH child is aware and feels inadequate to meet, one might posit two hypotheses: (a) In their estimates of their own inadequacy EMH children will be more derogatory of themselves than will be normal children of the same chronological age and grade placement. (b) Educable mentally retarded children in regular grades will be more derogatory of themselves than will the EMH children in special classes. In addition, the data were examined to determine whether the nature of self-derogation is different for each of the three groupings (normal, EMH in regular grades, EMH in special education) of children.

PROCEDURES

Sample

This study covers two essentially rural counties and a major industrial area in south-central Illinois. In the course of the large study, all first graders in specific districts in three counties were screened with the Primary Mental Abilities 5–7 test (N = 1807, mean age = 6.4). Those scoring 85 IQ or less were individually tested with the Peabody Picture Vocabulary Test and the Stanford-Binet, Form L. Of the 280 children individually tested, 120 were identified as having Binet IQ 60 to 85. Through randomization, 60 children were assigned to four special classes (two classes in an urban community and two in rural districts), and 60 children were left in the regular first grades wherein they were originally located. A third group of 60 children was selected from the remaining population to act as a "criterion" group of normal (PMA 95 to 110) first graders. The criterion group was matched to the EMH group on the basis of residence in the same community, by father's occupation, and family income.

Development and Description of Instrument

At the end of the first year of schooling all 180 subjects completed an instrument developed for this study, the Illinois Index of Self-Derogation (IISD).

Due to the paucity of empirical research upon the self-concept of young children, normal as well as retarded, it was deemed necessary to develop an instrument. In developing an instrument, the working definition adopted was "anything said about or attributed to one's self is, at least in part, one's self-concept." This definition not only describes a significant manifestation of self-concept, but also permits objective measurement. The existence of an unexpressed, unmeasurable matrix of components of the self-concept cannot be demonstrated, though few would deny its existence. However, there are responses which are put forth solely to influence favorably the behavior of others, because this is socially desirable; therefore these responses cannot be used to evaluate the subject's self-concept. These must be distinguished before this definition can be useful.

The IISD as employed in this study is designed for administration by the classroom teacher to groups of five children. While it was empirically determined that seven children of primary grade, age, or ability could complete the IISD at one time, five proved a more manageable group. The Index consists of 30 items, each item consisting of two sentences. One sentence described a child with a socially undesirable ascription, as determined by pretesting, such as: "Some children do not like the child with the balloon." The second sentence describes the child with neutral or socially desirable ascription, e.g., "Many children like the child with the flag." The child listens to the descriptions of the two children and then marks in a designated area of his score sheet the child which he regards as "most" like himself. The score sheet presents to the child three pairs of stick figures that have been empirically shown to be equally attractive to children.[1] The figure of the socially undesirable child is presented an equal number of times on the left and on the right. Of the 30 items, two are warm-up items, 22 are scorable derogations, and six are reliability check items.[2]

Given the choice between a socially desirable or acceptable response and a socially undesirable response, what are the possible interpretations which can be made when the child selects the undesirable response? The negative statement may be interpreted in two ways: Either this is the child's negative self-concept, or the child is deliberately making the unapproved response in this instrument. In either case, we define him as derogating himself from the point of view of his peers; and behavior, his own and others', will be based upon this negative self-concept. The IISD measured the tendency toward the specific nature of self-derogation.

Pretests

Two pretests were performed on nursery school children. The social desirability of items was determined by presenting the items to children between the ages of 2½ and 3½ and requesting that they label each child as "good" or "bad." The resultant instrument was then pretested in its present form and by the method described on 20 3½- to 4½-year-old children. Of this group only one child proved untestable.

The responses on the specific items had an average test-retest reliability of 80 per cent agreement. Almost all of the children had attended this nursery school in excess of one year. At a formal staffing each child was considered individually. The opinions of the staff as to the child's problems and attitudes toward self were compared with the results of the Index. The number of derogations ascribed by each child were defined as the extent of his negative self-concept. The specific ascriptions were interpreted literally. None of the items was found to be generally invalid. . . .

Factor Analysis

The IISD was factor analyzed. To aid in the description of factors, common technique is insertion of a "marker" variable; such an item would provide a simple and precise operational definition of a suspected factor. Since the research centered on unfavorableness of self-concept, the marker used in this analysis consisted of the absolute number of derogations made. In a principle axis solution this item had the highest loading (.962) on the first factor. It may then be said that the first and principal factor of the Index is degree of self-derogation. . . .

RESULTS

The first hypothesis (EMH children will be more derogatory of themselves than normal children) was verified through reference to the distribution of the number of derogations made by each group: The normal "criterion" group ascribed to themselves 2.6 derogations while the combined EMH group ascribed a mean of 3.2 derogations. The observed difference in distributions is significant at the .05 level ($Z = 5, p < .003$).

The second hypothesis (among EMH children, the control group, remaining in the regular grade, will be more derogatory of themselves

---

[1] In the nursery-school pretest, of 24 children questioned "Which child do you like best?" 11 preferred the "balloon-child" and 13 preferred the "flag-child."

[2] Protocols with less than 67 per cent internal consistency were judged invalid.

Table 1: *Proportion of each population ascribing each derogatory choice to themselves*

	Population		
	Normal "Criterion"	Experimental	Control
1. The mother of this child does not miss him when he is in school.	20	54*	52*
2. There are many things that this child does not know.	9†	8†	28
3. The mother and father of this child wish he would do better.	16†	35†	58
4. This child is going to have a hard time in school.	2	27*†	6
5. This child still won't be reading next year.	2	21*†	6
6. The mother of this child does not love him.	5	27*†	8
7. Kids like to make this child cry.	9	33*	18
8. People do not say nice things about this child.	9	33*	22
9. Some kids say nasty things about this child	11	35*†	10
10. This child cannot run as fast as other children.	27	41†	18
11. This child is not the same as other children.	25	43†	20
12. The mother and father of this child are proud of his brother or sister.	50	86*†	66

** Difference from normal group significant beyond .05 level.*
† Difference from control group significant beyond .05 level.

than the experimental special class group) was tested through reference to the distribution of the number of derogations made by each of the EMH groups. The control group had a mean of 3.0 derogations while the experimental group's mean was 3.4. The differences in distributions is significant at the .05 level but in the opposite of the hypothesized direction ($Z = -2.62$, $p < .004$).

A check was made of the proportion of each population ascribing each derogatory choice to themselves (*see* Table 1). The normal "criterion" group was found to have selected no specific items with greater than chance frequency. The entire EMH population was found to have selected item 1, "The mother of this child does not miss him when he is at school," with greater than chance frequency. There were two items ascribed by the EMH control group to themselves: 2, "There are many things that this child does not know"; 3, "The mother and the father of this child wish he could do better." The remaining nine items (4 to 12) of Table 1 were ascribed specifically by the experimental group to themselves.

It has been shown that the members of the groups ascribed derogations to themselves as determined by the group membership and not by random selection. . . .

DISCUSSION

This article has presented the data of the first year of a four-year project. As the project continues, further research emanating from it will be published.

The data support the first hypothesis. The results regarding the second hypothesis were significant but opposite to the predicted direction.

First, this study indicates that, even during the first year of school, significant differences can be shown between the self-concept of EMH children and normal children. Thus, the unawareness of or insensitivity to status usually attributed to primary age EMH children does not appear to be warranted. Second, it may be said that, after the first year of placement in the special class, EMH children of this study developed a more negative self-concept than they would have if left in their regular classes.

The specific derogations made by the EMH child have been shown to vary by the group to which he was assigned. If one assumes that, due to random assignment to the group, the two groups began the study with equivalent mean self-concepts, then the specific derogation made might be ascribed to the experimental condition.

Three possible explanations are offered for the second finding which is in contradiction to the bulk of previous research. Commonly, the EMH child placed in a special education class had already compiled a history of academic failure. This history of failure will determine the child's perception of himself and the attitude toward school and future progress therein. In this investigation, the impact of segregation from his peers was not modified by the alleviation of a potentially deleterious condition of future academic failure as was the case in all previous investigations. Therefore, the first possible explanation might be that the children have experienced segregation without realizing the compensation of improving conditions within the school situation.

The nature of the normal first grade curriculum places the emphasis upon rote learning. In later years, the emphasis of the normal curriculum shifts toward understanding. Therefore, the second possible explanation of finding 2 might be that the control children have not yet experienced the full impact of academic failure. Research indicates that EMH children can learn rote tasks with a minimum of additional difficulty but experience great difficulty in the application and generalization of their learning. Therefore, we would predict that the observed differences between the groups and the evaluation of themselves will shift to favor the experimental group as the curriculum of the control group shifts away from rote learning and the control group children experience a greater history of failure within the classroom situation. The experimental children, having developed a basis for generalization of learning and having had understanding emphasized over rote learning in this first year, will then show greater overall improvement in the area of academic confidence and this will be reflected in their improved self-concepts.

A third possible explanation is suggested by a review of the derogations which the experimental group made. These children may realize that they (a) are less competent in motor skills and interpersonal relations, (b) have difficulty in learning to read, and (c) will experience difficulty throughout their academic career. This does not obviate the fact that the experimental group made a larger number of derogations, but it does allow for the possibility that in the following years of the project this reality orientation may become the basis for a positive self-concept founded upon the evaluation of what the individual can achieve rather than his failings. The latter is a critical factor if there is any merit in the assumption that reality based self-concept is crucial to a healthy personality (Rogers, 1951). Therefore, it is suggested, on the basis of the derogations which the experimental group ascribed to themselves, that at the conclusion of the entire five-year study special class placement may be preferable to nonplacement for EMH children.

With the exception of those having a psychoanalytic bent, sociological and psychological investigators have hesitated to investigate the self-concept of young children. It is suggested that some of the implications of this article are that:

1. The roots of social expectations and acceptable behavior can be tapped as early as 2½ to 3½ years of age.
2. A consistent self-concept can be measured in children as young as 3½ to 4½ years of age.
3. The nature of the deviation called mental retardation affects the self-concept even when it is not severe (Binet 65 to 85) and has not been given official recognition, and thus can be measured at age 7.
4. When given official recognition by segregation, mentally retarded children, as described above, develop much clearer and more negative self-concepts.

SUMMARY

Of the 1807 entering first grade children, the total registration of three areas, 120 were identified as having Binet IQs of 60 to 85. Half of these children were randomly assigned to special classes and half to regular classes. An additional "criterion" group of 60 normal children were identified to match the retarded sample with regard to area of residence, father's occupation, and family income. These 180 children were tested with the Illinois Index of Self-Derogation, developed for this investigation, at the end of their first year of schooling. The results indicate significant differences in the mean number of derogations made both between the retarded group and the normal "criterion" group, favoring the criterion group, and between the control and experimental group, favoring the control group. Ascription of specific derogations was shown to be relevant to group membership. Factor analysis of the data verified that the Index has as its primary factor self-derogation.

REFERENCES

JORDON, T. E., & DECHARMES, R. The achievement motive in normal and mentally retarded children. *Amer. J. ment. Def.*, 1950, *55*, 60–89.

ROGERS, C. *Client-centered therapy*. Boston: Houghton Mifflin, 1951.

PART IV

MIDDLE CHILDHOOD

PARENT-CHILD RELATIONS

THE CHILD IN SCHOOL

PEER RELATIONS

COGNITIVE DEVELOPMENT

PERSONALITY DEVELOPMENT

THE IMPACT OF THE CULTURE

PARENT-CHILD RELATIONS

28

Effects of Maternal Employment on the Child

Lois Wladis Hoffman

Numerous articles have been written in both scientific and lay journals about the working mother and the effects her employment has on the child. Hoffman's study illustrates the complexity of the problem and the need to get beyond "working" as the major variable to be analyzed. This study again demonstrates the use of research to test particular theories arrived at from literature in the field.

Empirical studies of the effects of maternal employment have long suffered from a paucity of adequate controls. It was the lack of controls, for example, which led to the long-standing belief that maternal employment was strongly associated with juvenile delinquency. In fact, both maternal employment and delinquency were associated with social class and with broken homes; and, when the latter variables were controlled the relationship disappeared, suggesting it had been a spurious one (Glueck & Glueck, 1957).

Yet, the new studies on maternal employment, despite their more adequate controls, have often produced more confusion than clarification. Maternal employment has been associated with *less* delinquency, *more* delinquency, withdrawal behavior, dominance behavior, and nothing at all. Furthermore, there are existing theories to account for each of these findings. Thus, delinquent and hostile-assertive behavior might result from parental neglect and lack of sufficient discipline due to the mother's absence from the home; from hostility on the part of the child at being deserted; or, in conjunction with Henry's notion (1957) that father discipline leads to the expression of hostility outward whereas mother discipline leads to intrapunitiveness, from father discipline replacing mother discipline.

Withdrawal and dependency responses by the child might also be expected. Loss of mother through death or hospitalization has been associated with overdependency—the young child perhaps seeking in this crisis a return of the maternal relationship he had in the past. Something similar might occur as the result of the mother's spending a great deal of time outside the home working. Still another reason for withdrawal and dependency might be that the working mother feels guilty about working outside the home and responds with a pattern of "smother love" or overprotection. Such a pattern could alleviate the mother's guilt. In addition, if it is true that working mothers have greater power needs than nonworking mothers, it would also be an appropriate pattern for expressing power over the child in a way that appears to be benign and admirable.

Some, such as Cyrus (1952), would expect

From *Child Development*, 1961, *32*: 187–197. Reprinted with permission of the Society for Research in Child Development.

working mothers to be less frustrated and more capable of warm interaction. Others would say that the energy output required by trying to fill two jobs would make the mothers more irritable with their children.

Instead of speculating as to which of these theories is correct, the present paper will suggest and illustrate an approach in which the question is: *When does one pattern operate and when does another?*

Each theory assumes something about the working situation that is sometimes, but not always, correct; and, when certain assumptions hold, different mother-child interactions will result and the child will be affected accordingly. One important factor implicit in several of the theories involves the question of whether or not the mother enjoys her employment. It is this difference that the present study has selected for focus. It was expected that the effects of employment on the mother-child relationship, and consequently on the child, would be different when the working mother enjoyed work than when she did not.

Two of the theories mentioned above, the "guilt-overprotection" theory and the "neglect" theory, lead to very different predictions about the mother's behavior toward the child, and, although both theories predict that the working mother's child will be disturbed, the pattern of disturbance predicted by each is very different. Both these theories served as guides in the present investigation. The "guilt-overprotection" theory was expected to operate for the working women who enjoyed work, and the "neglect" theory for the working women who did not enjoy work.

Only the working woman who enjoys her employment should be guilty about it, and as such it was predicted that such a woman would try to compensate for her employment by showing a great deal of affection toward the child, disciplining the child adequately to leniently, and being careful that the child should not be inconvenienced by the fact that she works, e.g., by having to help with household tasks. The child correspondingly should be relatively nonhostile, nonassertive, and, if this parental pattern were carried to the point of overprotection, somewhat withdrawn and passive. The working woman who dislikes her work should, on the other hand, be quite guiltless. As such, she was expected to show more of a withdrawal from the maternal role—showing less positive affect and less disciplining. She should feel freer to inconvenience the child—specifically to expect the child to help with household tasks. The child, feeling somewhat resentful and lacking sufficient discipline, was expected to be assertive and hostile. The predictions were, then, that both groups would show different patterns than when the mother was not employed, but that these differences would not be the same.

PROCEDURE

The sample included 176 white, *intact* families with at least one child in the third through sixth grades of three elementary schools in Detroit. There were 88 working-mother families and 88 nonworking-mother families matched on occupation of father, sex of child, and ordinal position, including the important status of being an only child. Except for sex of child, all matched variables were selected because they were known to be related to maternal employment (Hoffman, 1960) and because it was believed they might be related to the dependent variables. Pairs of families were matched by sex of child so that it might be possible to examine the relationship between maternal employment and the dependent variables separately for boys and girls, to be certain that combining was legitimate. All statistical comparisons are between working and nonworking populations, and with a few exceptions which will be pointed out later, *all comparisons are for matched pairs.*

The data reported here are based on questionnaires filled out by the children, interviews with the mothers, teacher ratings, and a classroom sociometric. Each of these measures will be taken up in turn when the findings are presented.

In the interview with the mothers, the working mothers were asked, "How to you feel about working?" These answers were coded as to the predominance of positive or negative attitudes toward work. Responses to this question were on the positive side for most respondents. Sixty-five respondents were classified as liking work, and 23 as disliking it.

All hypotheses were tested separately for the two groups, i.e., by comparing the 65 mothers who liked work with their 65 matched nonworking counterparts and the 23 mothers who disliked work with their 23 matched nonworking counterparts. Significance tests for differences included the t test for correlated means where the data were quantitative, sign tests where the data were qualitative, and the t test for proportions where the data could not be analyzed by pairs.

Table 1: *Mother's affect and behavior toward child: working-nonworking comparisons by attitude toward work*

Variable	Working mothers with positive attitude toward work	Working mothers with negative attitude toward work
Child's report		
positive affect from mother	more than nonworking*	less than nonworking
coerciveness from mother	less than nonworking	less than nonworking*
Mother's report		
severity of discipline†	less than nonworking*	no difference
power assertion by the mother†	less than nonworking	less than nonworking*
mother's feelings of sympathy†	more than nonworking*	no difference
mother's feelings of hostility†	less than nonworking*	less than nonworking

Note.—Findings reported in all tables are based on the t *test for significance of difference between correlated means except where indicated.*

** Significant at the .05 level, one-tailed test.*

† Comparisons not possible between matched pairs. Each working group therefore compared to pooled sample of nonworking mothers and the t *test for significance of difference between proportions used.*

RESULTS

Table 1 summarizes the results dealing with the mother's affect and behavior toward the child. The first comparison deals with the child's perception of the mother as a source of positive affect. The child had been asked to indicate the persons who, from his own experience, best fitted each of a series of verbs, such as "praises" and "smiles," and to indicate the degree of fitness of person to verb. The perception of positive affect in the mother-child relationship was measured by the extent to which the child associated "mother" with the following words: "praises," "smiles," "listens," "helps," "explains." As can be seen in Table 1, where the working mother indicates a positive attitude toward her work, the child associates more positive affect with the mother than do children in the matched group of nonworking women. This difference is significant. When we compare the children of working mothers who have negative attitudes toward their work to the nonworking group, we find the opposite relation although it is not a significant one.[1]

The second set of findings deals with the child's perception of coerciveness on the part of the mother. This variable was similarly measured using the words, "threatens," "punishes," and "hits." The table shows that the working mother is less likely than the nonworking mother to be associated with coerciveness in both groups and that the relation is significant for the working mothers who dislike work.

The remaining findings in this table are based on the mother's report. In the interview, mothers were asked to give detailed accounts of two recent occasions when they wanted the child to do something that the child did not want to do. These descriptive accounts were coded for degree of severity of discipline, degree of power assertion over the child, sympathy, and hostility. Unfortunately, these descriptions were not available on all of the matched nonworking respondents. To have a large enough *N*, it was therefore necessary to pool the nonworking respondents and compare the two groups of working women separately with this pooled group. These are not, then, matched comparisons, although the working and nonworking groups are, on the whole, comparable with respect to the matching variables. The findings show that where the mother has positive attitudes toward employment she is less severe in her discipline, uses less power-assertive influence techniques, and feels more sympathy and less hostility during the interaction than is the case with the nonworking group. For the working mothers with negative attitudes toward employment there is less power assertion and somewhat less hostility.

Taking the findings in Table 1 as a whole, they tend to support the hypothesis that the working woman who gains pleasure from her work shows the child more affection and uses milder discipline. Although she is emotionally involved in the discipline situation, the involve-

---

[1] The direction of the relation is reported even where differences are small; so, except for the significant relations, differences can be due to chance variations.

Table 2: *Child's participation in household tasks: working-nonworking comparisons by attitude toward work*

Variable	*Working mothers with positive attitude toward work*	*Working mothers with negative attitude toward work*
Child's report		
regular participation in household tasks	less than nonworking	more than nonworking*

** Significant at the .05 level, one-tailed test.*

Table 3: *Child's behavior: working-nonworking comparisons by attitude toward work*

Variable	*Working mothers with positive attitude toward work*	*Working mothers with negative attitude toward work*
Mother's report		
child's assertiveness toward mother‡	no difference	more than nonworking *
child's playing with younger children rather than own age or older§	more than nonworking*	no difference
Teacher's report		
initiation of friendships	less than nonworking*	less than nonworking
influence attempts made to peers	less than nonworking	more than nonworking
influence success with peers	less than nonworking*	no difference
teacher dependency (girls)	no difference	no difference
teacher dependency (boys)	more than nonworking	more than nonworking†
aggressiveness	no difference	more than nonworking
use of physical force	less than nonworking	more than nonworking*
impulse control	no difference	less than nonworking*

** Significant at the .05 level, one-tailed test.*
† Significant at the .05 level, two-tailed test, difference not predicted.
‡ Comparisons not possible between matched pairs. Each working group therefore compared to pooled sample of nonworking mothers and the t *test for significance of difference between proportions used.*
§ Sign test used for this variable.

Table 4: *Child's adjustment: working-nonworking comparisons by attitude toward work*

Variable	*Working mothers with positive attitude toward work*	*Working mothers with negative attitude toward work*
Teacher's report		
nonadaptive response to frustration†	more than nonworking	more than nonworking*
intellectual performance	less than nonworking*	less than nonworking*
Classroom sociometric		
liking by others	less than nonworking	less than nonworking

** Significant at the .05 level, one-tailed test.*
† Sign test used for this variable.

ment is to a great extent one of sympathy for the child. The working mother who does not enjoy her work, and whom we would expect to be relatively guilt-free thereby, also shows a tendency toward mild discipline. However, here it is not part of a larger pattern which includes affection and sympathy toward the child. Instead it seems to be part of a general pattern of less involvement with the child. This suggests that the dislike-work group tends to withdraw from the maternal role.

Table 2 deals with the child's participation in tasks. This was measured by the child's responses to a form of the "Doing Things" questionnaire developed by Herbst (1953), in which the child is asked to report the extent of his participation in several household tasks (20 tasks in this case). As expected, the children of the working mothers who do not enjoy their work participate more in household tasks than do the children of the nonworking mothers; the children of the working mothers who enjoy work participate less. This is in keeping with the notion that the mothers who enjoy work go out of their way to avoid inconveniencing their children because of their employment.

Table 3 deals with the child's behavior. The child's assertiveness toward the mother is based on the mother's account of the conflict situations in which she wanted the child to do something that the child did not want to do. The child's behavior was considered assertive when he directly indicated that he did not wish to comply and that he did not expect to do so. These behaviors included requests such as "Is it all right to stay up until the end of this program?"—as well as statements of intention such as "No. I am not going to bed until the end of this program."—where the context made it clear that the child was not "just talking." Pleading, crying, complying with complaints or silent anger, and devious attempts to obtain ends ("I have a stomach ache.") were all considered less assertive. The child's tendency to play with younger children is based on the mother's report that his playmates were "mostly older," "mostly younger," or "the same age." The remaining variables reported in Table 3 are based on teacher ratings. These ratings are described more fully elsewhere (Hoffman, Rosen, & Lippitt, 1960).

It was hypothesized that the children of working mothers who have positive attitudes toward employment would be nonhostile, nonassertive, and possibly withdrawn and passive as compared with their nonworking counterparts, but that the children of working women who do not like work would be assertive and hostile. The data seem to bear this out. The children whose mothers like work differ in their general lack of initiative-taking and tend to play with children younger than themselves. The children whose mothers do not like work, on the other hand, show assertive behavior toward their mothers and toward their peers. This assertiveness toward peers includes aggressiveness, use of physical force, and less impulse control in general. Only teacher dependency does not operate as expected, the boys in both groups showing more dependency, significantly more in the dislike-work group.

Both the "guilt-overprotection" theory and the "neglect" theory lead to the prediction that the children of working mothers will be more disturbed in general than the children of nonworking mothers. This was tested by comparing the matched pairs on the use of nonadaptive responses to frustration, intellectual performance, and sociometric ratings by classmates. The results of these tests are reported in Table 4. The first measure was based on a teacher rating of how the child typically reacts to frustration. "Realistic acceptance of insolvable problem" and "attacks problem directly" were considered adaptive responses, while responses such as "cries," "goes to pieces," "blames others," and "becomes overcritical of self" were considered nonadaptive. Both groups of working-mother children were more likely to use nonadaptive responses to frustration. They also showed lower intellectual performance and were somewhat less liked by the other children in the class.

The lower intellectual performance of children whose mothers work is a particularly interesting finding. Being low on performance may be a function of low ability or low motivation. If it is true that the working mothers who like work are guilty about their employment and consequently overprotect their children, then the intellectual *ability* of their children may be impaired. Mothers who solve their children's problems for them may hamper their intellectual development by depriving them of valuable problem-solving experience. Thus, these children would be expected to be low on ability as well as performance. The low performance of the dislike-work children, on the other hand, might be part of the rebellious pattern that seems to characterize this group; here we would not necessarily expect low ability. Scores on the Primary Learning Aptitude Test for third graders and the Detroit Alpha Intelligence Test for fourth, fifth, and

Table 5: *Occupation and attitude toward work*

Occupation	Working mothers with positive attitude toward work	Working mothers with negative attitude toward work
Professional and semiprofessional, e.g., teacher, social worker, registered nurse	14	1
Skilled, e.g., secretary, bookkeeper	11	7
Sales	14	0
Waitress	1	4
Semiskilled or unskilled factory and clerical	20	10
Domestic	5	1
Total	65	23

sixth graders were available. The expectations were borne out. Only the children whose mothers were positive about their work obtained lower scores on these tests than their nonworking counterparts, and this difference was significant. The others showed no such difference.

SUMMARY AND CONCLUSIONS

The over-all pattern of findings suggests that the working mother who likes working is relatively high on positive affect toward the child, uses mild discipline, and tends to avoid inconveniencing the child with household tasks; the child is relative nonassertive and ineffective. The working mother who dislikes working, on the other hand, seems less involved with the child altogether and obtains the child's help with tasks; the child is assertive and hostile.

The data do not show conclusively that the reason for these different patterns is the presence or absence of guilt in the mother, but they are consistent with such an interpretation. If one were to argue that the causality is reversed, that the attitude toward work is a function of the effect of employment on the child or that happier women produce happier children, it would not be consistent with the fact that both groups of working-mother children show indications of maladjustment. It is only that they show different *syndromes* of maladjustment. Furthermore, in Table 5 it can be seen that the attitude toward employment is in large part a function of the particular job the woman holds.

It is interesting to note in Table 5 that, while attitude toward employment is not purely a function of social class, there is somewhat of a tendency for higher status jobs to be well liked. Because middle-class women, compared to lower-class, can obtain higher status jobs, because their employment is likely to be choice rather than necessity, and because there is probably greater anxiety about childrearing in the middle class, it is likely that guilt and overprotection are more a middle-class response to maternal employment while neglect is more a lower-class response.

The real purpose of this paper, however, is not so much to elaborate on any particular theory as to suggest an approach to the study of the effects of maternal employment. The need to control on spurious variables has been established. The problem now is to differentiate the all-inclusive variable of maternal employment into more highly specified categories. Only in this way can we come to understand the processes by means of which employment affects the child. If one wants to know whether or not the sheer absence of the mother from the home is the important variable, then one should subdivide the working women according to the number of hours they are away from home while the child is there. If one wants to know the importance of the adequacy of supervision, then one should subdivide the women according to whether or not the substitute supervision provided is adequate. One might feel that the important variable is the father's reaction to the mother's employment. In this case, the working women can be subdivided according to some measure of the father's attitudinal or behavioral response to his wife's employment. In short, by selecting a particular aspect of mother's employment, differentiating the working mothers according to this variable, and examining the differences between working and nonworking groups separately, one can make a more formidable step toward understanding the crucial aspects of maternal employment in relation to the child.

There are still other advantages of this kind of approach. When the working-mother group is not differentiated, it includes a heterogeneous group and, as such, the working-nonworking comparisons will obscure differences due to counteracting subgroup relations. For example, in the data reported here, since one group of working-mother children was less assertive and the other group was more assertive, the two tendencies canceled one another out. Thus, when the total group of working-mother children was compared with the matched nonworking group, there appeared to be no differences in child assertiveness. The same is true for the participation of children in household tasks.

Another kind of error that can result from failure to differentiate the working mothers is when what seems to be a single pattern of findings actually includes two or more distinct patterns. For example, when the total groups were compared, working mothers' children were found to play more with younger children and also to show less impulse control. Doing separate analyses for the two groups of working mothers, however, revealed that these results did not come from the same subjects. Although it was not true of the present data to any great extent, it would be possible for a pattern to emerge from the general working-nonworking comparisons which actually characterized neither the total group nor any single subgroup, but was entirely a coincidence of findings contributed by different clusters of subjects.

One final point should be made. In most working-mother studies, relations are examined between maternal employment and some aspect of the child's behavior *directly,* without considering intervening variables such as the effect of maternal employment on parent-child relations, family structure, or other aspects of family life. The jump between maternal employment and child behavior is too broad to be covered in one leap. One must first understand the various steps that intervene.

In summary, this paper suggests that maternal employment has a different effect on the mother-child relationship and on the child's behavior depending on whether or not the mother enjoys working. The purpose in doing this analysis was to show that greater understanding of the effects of maternal employment will come about when working mothers are differentiated along some theoretically relevant dimension and when the effects on family life, as well as on the child, are examined.

REFERENCES

BANDURA, A., & WALTERS, R. H. *Adolescent aggression.* New York: Ronald Press, 1959.

CYRUS, D. Problems of the modern homemaker-mother. In J. T. Landis & M. G. Landis (Eds.), *Readings in marriage and the family.* New York: Prentice-Hall, 1952. Pp. 392–402.

GLUECK, S., & GLUECK, E. Working mothers and delinquency. *Ment. Hyg.,* 1957, 327–352.

HENRY, A. F. Sibling structure and perception of the disciplinary roles of parents. *Sociometry,* 1957, *20,* 67–74.

HERBST, P. G. Analysis and measurement of a situation: The child in the family. *Hum. Rel.,* 1953, *6,* 113–140.

HOFFMAN, LOIS W. Effects of the employment of mothers on parental power relations and the division of household tasks. *Marriage Fam. Living,* 1960, *22,* 27–35.

HOFFMAN, LOIS W., ROSEN, S., & LIPPITT, R. Parental coerciveness, child autonomy, and child's role at school. *Sociometry,* 1960, *23,* 15–22.

29

Verbally Expressed Needs and Overt Maternal Behaviors

Vaughn J. Crandall and Anne Preston

This study by Crandall and Preston is part of a long-range group of studies of the Fels Research Institute. An important issue in the introduction of this book was the problem of self-report versus observation in the study of parent-child relationships. This study attempts to correlate the mother's needs as expressed in a personality inventory with her observed behavior. The results indicate that such a correlation does exist, at least in the areas of aggression and nurturance.

From *Child Development,* 1961, *32:* 261–270. Reprinted with permission of the Society for Research in Child Development.

Research concerned with parent-child relationships and the influence of parents on children's social learning and personality development has gradually elucidated a number of avenues through which parents affect their children's development. Recently, child personality development research has been primarily focused on two major avenues of parental influence—the effects of direct parental reinforcements and the more amorphous impact of parents as learning models for children's imitation and identification. Knowledge of the processes of parental influence is, however, incomplete until parents are also studied as individuals *per se,* as well as parents *qua* parents, and relations are explored between the parents' *own* needs, aspirations, and defenses and the behavior they exhibit with their children.

The present study investigates relations between expressed needs of a group of middle-class women and specific overt behaviors they display toward their children. The investigation focuses on the general question of how, and to what degree, such needs find overt expression in maternal behaviors. The study, in contradistinction to most recent research on parental influence relying solely on *reported* parental behaviors obtained through interviews or questionnaires, is based on data of *observed* maternal behaviors of the *S*s as they interacted with their children in their homes.

METHOD

Subjects

The sample was comprised of 38 white, middle-class women, mothers of families participating in the Longitudinal Study of Child Development of the Fels Research Institute for the Study of Human Development and living in industrial centers, in small towns, and on farms within a 30-mile radius of the Institute in southwestern Ohio.

Assessment Procedures

Assessment contacts were made with the *S*s at the Fels Research Institute and in the *S*s' homes. At the Institute the *S*s were administered, among other tests, the Edwards Personal Preference Schedule (PPS). Their responses to the PPS were scored in accordance with the PPS manual (Edwards, 1954), and the scores for the following needs were used as part of the data of the investigation: *n* aggression, *n* dominance, *n* nurturance, and *n* autonomy. These needs were selected for study as ones which might be especially likely to enter into, and influence, maternal behaviors.

Maternal behavior data were obtained through the Fels Home Visit Program. This program, a part of the regular longitudinal data collection procedures of the Fels Longitudinal Study, provides for periodic observations of mother-child interactions as they occur in the homes of the *S*s. A home visitor, an experienced clinical psychologist specially-trained for the Home Visit Program, makes at least two observation-visits to the homes of the *S*s each year. Subsequent to each of these visits, the home visitor rates the *S*'s maternal behaviors using the 30 Fels Parent Behavior Rating Scales (PBRs). When observations cover a mother's interactions with more than one of her children, separate PBR ratings are made in respect to the mother's behavior with each child. For the present study, when PBRs were obtained regarding more than one child in a family, the PBR ratings of the mother's behaviors with *one* of her children were randomly selected; the PBR data for the other child or children were excluded. The PBRs so selected represented ratings of maternal reactions towards children ranging from 3 through 8 years of age. The investigation of maternal reactions toward children of such a broad range of ages might be questioned; certain maternal behavior could differ significantly depending on the age of the child. However, previous research (Baldwin, Kalhorn, & Breese, 1949) has found negligible differences in PBR ratings as a function of age of child within the age span used in the present study.

The time-consuming nature of the Home Visit Program precluded the possibility of determining PBR interrater reliabilities for the sample of this specific study. However, previous research based on the Fels Home Visit Program and the Parent Behavior Rating Scales has demonstrated acceptable inter- and intrarater reliabilities for the scales employed in the current study.

The PBR ratings of observed maternal behaviors used in this study were those obtained during the home visit closest in time to the occasion when each *S* was administered the Edwards PPS at the Fels Institute. The median time elapsing between home visit observations and PPS administrations was five weeks.

Hypotheses

The 30 Fels Parent Behavior Rating Scales

Table 1: *PBR scales representing the three major dimensions of maternal behavior investigated in the present study*

PBR Scales Representing Affectional Maternal Behavior

Affectionateness toward child: hostile—affectionate
Rapport with child: distant—close
Direction of criticism: disapproval—approval
Intensity of contact: vigorous—inert

PBR Scales Representing Protective Maternal Behavior

Protectiveness: exposing—sheltering
Solicitousness for child's welfare: nonchalant—anxious
Babying: withholds help—overhelps
Child-centeredness of home: child-centered—child-subordinated

PBR Scales Representing Coercive Maternal Control

Severity of penalties: mild—severe
Readiness of enforcement: lax—vigilant
Coerciveness of suggestions: compliance is optional—mandatory

were originally constructed to assess a large range and diverse variety of parental behaviors (Champney, 1941). Data obtained with these scales have been subsequently analyzed for major "underlying" dimensions of parent behavior by a number of statistical techniques including syndrome analysis, cluster analysis, and first- and second-order factor analyses (Baldwin *et al.*, 1949; Crandall & Preston, 1955; Lorr & Jenkins, 1953; Roff, 1949). These various analyses have identified three—and sometimes four, depending on method of analysis—major dimensions of parent behavior. . . . Three of these dimensions were selected for investigation in the present study and are listed in Table 1. The intercorrelations of the variables representing these three dimensions range from .60 to .84 and indicate that, while the variables within each cluster intercorrelate highly enough to justify the assumption of an underlying dimension, the correlations within each cluster were not sufficiently high to permit the PBR scales of these variables to be used interchangeably. In none of the correlations was as much as three-quarters of the variance held in common. In the present study, therefore, while data analysis was organized around the three major maternal behavior dimensions investigated, and general hypotheses were stated in respect to these, tests of these general hypotheses were made by ascertaining relations between the *S*s' PPS need scores and *each* of the PBR variables representing the maternal behavior dimensions under investigation.

The study explored relations between middle-class women's verbally-expressed needs and their overt maternal behaviors. Operationally, the *S*s' needs were determined by their Edwards PPS responses; their maternal behaviors were quantified in PBR ratings based on home visit observations of mother-child interactions.

Four general hypotheses were tested. These related the *S*s' *n* nurturance, *n* aggression, *n* autonomy, and *n* dominance to maternal behavior dimensions. *Hypothesis I:* "Middle-class women's *n* aggression will be negatively related to the amount of overt affection they display toward their children." Parenthetically, it should be noted that hypothesis I, as well as the three hypotheses which follow, is tenable only to the degree that women's verbally-expressed needs find direct expression in maternal behaviors. When internal or external inhibiting forces are present, such needs may not eventuate in relevant actions. Women with aggressive propensities, for example, may not behave aggressively toward their children when internal checks are present such as the anticipation of guilt following aggression and/or when external inhibiting forces exist, e.g., possible criticism from others for hostile maternal behavior. In addition, mothers may report aggressive urges but use persons other than their children as primary targets for their aggression, e.g., husband, acquaintances, or various minority groups. Nevertheless, while some or all of these possibilities may occur, the present study tests the general proposition that middle-class women's verbally-expressed needs will, beyond chance, find behavioral expression in their parent-child interactions. In reference to hypothesis I, it is expected that women who freely acknowledge aggressive urges will express such aggression overtly toward their children by nonaffectionate and rejectant behaviors. Concerning the specific PBR variables representing the dimension of maternal affection (cf. Table I), it is predicted that, in comparison with *S*s with low *n* aggression, *S*s who readily report aggressive desires will: (a) express less overt affection toward their children, (b) maintain less close rapport with them, (c) tend generally to criticize rather than approve of their children's behaviors, and (d) have less intense contacts with their children.

Hypotheses II and III are concerned with women's needs as predictors of their overt protective maternal behaviors. *Hypothesis II:* "Middle-class women's verbally-expressed *n* nurturance will be positively related to the amount of protectiveness they display toward their children." In respect to specific protective maternal behaviors (cf. Table 1), it is predicted that *S*s with high *n* nurturance will be especially prone to: (a) act protectively toward their children, tending to shelter them from, rather than exposing them to, potential physical and psychological vicissitudes, (b) be solicitous, i.e., concerned rather than nonchalant, regarding their children's welfare, (c) infantalize their children by overhelping rather than withholding help, and (d) maintain a home in which activities are child-centered rather than child-subordinated.

Hypothesis III relates women's *n* autonomy and their protective maternal behaviors. *Hypothesis III:* "Middle-class women who express strong needs for autonomy will exhibit little protective behavior toward their children." The rationale underlying this hypothesis is less straightforward than those of hypotheses I and II. For the two previous hypotheses, it was assumed that the needs investigated would be directly translated into maternal behaviors; *n* aggression would result in rejectant maternal behaviors and *n* nurturance, in protective maternal behaviors. The reasoning behind hypothesis III is more circuitous. In fact, either of two processes may produce the predicted relation between *n* autonomy and maternal protectiveness. First, mothers with strong needs for autonomy and independence might be expected to value similar needs in their children—at least to the degree that these mothers identify with their children and react to them as extensions of themselves. Such mothers might, then, instigate and reward autonomous behaviors of their children and discourage dependent overtures such as help-seeking or emotional support-seeking calling for maternal nurturance.

A second possibility also exists. Mothers with strong *n* autonomy may *not* necessarily value autonomous behavior of their children and, therefore, react nonnurturantly when their children make dependent demands on them. Nevertheless, such women may still exhibit little nurturance with their children insofar as they feel their children's dependent overtures are encroachments on their *own* need for autonomy. If the latter were the case, these women might be expected to be especially likely to have child-subordinating, rather than child-centered, orientations (cf. Table 1). In general, however, in respect to hypothesis III, it is predicted that women with a strong *n* autonomy will, in comparison with women with low *n* autonomy, be: (a) less protective of their children, (b) less babying, (c) less solicitous, and (d) less child-centered.

A final hypothesis relates *n* dominance and coercive maternal controls. *Hypothesis IV:* "Women who readily express a need to dominate others will be especially prone to employ coercive controls with their children." Specifically, it is predicted that *S*s with high *n* dominance, in comparison with *S*s with low *n* dominance, will: (a) evoke more severe penalties when their children do not follow their rules and regulations, (b) more readily enforce the rules they establish for their children's behavior, and (c) use more coercive suggestions with their children regarding what they—the children—should or should not do.

RESULTS

Table 2 presents relations between the *S*s' verbally-expressed needs and their overt maternal behaviors. The data in this table have been organized in reference to the four major hypotheses of the study as well as in respect to the specific PBR variables representing the three general maternal behavior dimensions investigated.

Hypothesis I, that the *S*s' *n* aggression would be negatively associated with the amount of affection they displayed with their children, found general confirmation. Women with high PPS *n* aggression expressed less overt affection toward their children, had less close rapport with them, tended to criticize rather than approve of their children's behavior, and had less intense contacts with their children. In general, then, as represented by the sample of the present study, the verbally-expressed need for aggression of middle-class women *does* find direct behavioral expression in the nature of the affectional rela-

Table 2: *Relations between subjects' verbally-expressed needs and their observed maternal behaviors* ($N = 38$)

PPS Variable PBR Variable (and direction)	*r*	*p*
Hypothesis I—Aggression		
Ready expression of affection	−.40	.01
Close rapport with child	−.53	.01
Direction of criticism (approval)	−.54	.01
Intense contact with child	−.36	.05
Hypothesis II—Nurturance		
Much protectiveness	.48	.01
High solicitousness	.50	.01
Tendency to baby child	.44	.01
Child-centered orientation	.45	.01
Hypothesis III—Autonomy		
Much protectiveness	−.25	*ns*
High solicitousness	−.22	*ns*
Tendency to baby child	−.23	*ns*
Child-centered orientation	−.39	.05
Hypothesis IV—Dominance		
Employs severe penalties	.23	*ns*
Readily enforces rules	.29	*ns*
Coercive suggestions	.07	*ns*

tionships these women establish with their children.

Hypothesis II, that the *S*s' *n* nurturance would be positively related to the amount of protectiveness they expressed toward their children, also found substantiation. The *S*s whose Personal Preference Schedule responses indicated a strong general need to nurture others were especially protective of their children. In comparison with the *S*s with low *n* nurturance, in everyday life these women exhibited more protectiveness toward their children, were more solicitous of their children's welfare, tended to overhelp rather than underhelp them, and established homes where child-centered orientations predominated.

Hypotheses III and IV received less clear-cut support. In all four correlations concerned with hypothesis III, directions of relations were those predicted; *n* autonomy tended to be negatively associated with protective maternal behaviors. However, only one correlation—that between *n* autonomy and child-centeredness—was statistically significant. All relations concerned with hypothesis IV were in the predicted direction; none was significant.

DISCUSSION

The results indicate that some verbally-expressed needs of middle-class women are directly translated into maternal behaviors. Verbal reports of aggressive urges significantly predict a number of different rejection maternal behaviors. Similarly, the expression of a general desire to nurture others found behavioral expression in the *S*s' everyday interactions with their children. However, other needs were not, as hypothesized, significant predictors of the *S*s' behavior with their children. Only one of four correlations between the *S*s' *n* autonomy and their protective maternal behaviors was statistically significant. The fact that the significant relation occurred between high *n* autonomy and low child-centeredness suggests—though does not conclusively prove—that the need for autonomy in middle-class mothers is less a determinant of low maternal protectiveness as a way of developing valued autonomous behavior in their children than it is a more direct reflection of egocentric desires by the strong *n* autonomy mothers to be "left alone" by their children. It appears that high *n* autonomy mothers may express less nurturant maternal behaviors simply because they feel their children's dependent demands are encroachments on their *own* autonomy.

The final hypothesis that the *S*s' needs to dominate others would find behavioral expression in the coerciveness of the control methods they employed with their children was not substantiated. Several alternative *post hoc* explanations can be given for the lack of significant relations between variables testing hypothesis IV. None of these explanations, however, can be tested with the data of the present study. One possibility is that the *n* dominance scale of the Edwards PPS is not a predictively useful measure of adult women's need to dominate others (including their children). The PPS, while it is one of the most meticulously constructed of current personality questionnaires (e.g., controlling for social desirability of responses), still awaits construct validation research. In a limited sense, the current study presents evidence for the construct validity of the PPS *n* aggression and *n* nurturance scales. The scales *do* predict the amount of aggression and nurturance mothers express toward their children. It is also interesting to note that middle-class mothers' PPS *n* aggression and *n* nurturance scores are as good, or better, predictors of their overt affectional and nurturant maternal behaviors as are their own self-ratings of their maternal behaviors in these areas. In a previous study (1955), the authors constructed self-rating scales of maternal behavior to parallel the same Fels PBR scales employed in the present study. In the previous study, a group of middle-class mothers similar to, and including some of, the *S*s of the present study rated their own maternal behaviors with these specially-constructed mother self-rating scales. These self-ratings were correlated with parallel PBR ratings made by the Fels Home Visitor on the basis of her observations of these mothers' overt interactions with their children. In respect to the general maternal behavior dimension of affection, correlations between the mothers' self-ratings and the Home Visitor's PBR ratings were .59, .33, and .45 for affectionateness, rapport, and direction of criticism, respectively. These correlations are of the same general magnitude as those obtained in the present study between PPS *n* aggression scores of the *S*s and these same PBR variables (cf. Table 2). In other words, middle-class women's PPS *n* aggression (an assessment of need not specific to mother-child relationships) predicts their overt affectional maternal behaviors as well as do the mother's own self-ratings of these specific maternal behaviors. Moreover, in respect to the prediction of protective maternal behaviors, women's PPS *n* nurturance scores (again not specifically pertaining to mother-child relationships) predict overt protective maternal behaviors *better* than do the mothers' own self-ratings of these behaviors. In the present study, PPS *n* nurturance was significantly correlated with PBR ratings of protectiveness, babying, solicitousness, and child-centeredness; in the previous study, mothers' self-ratings of these same behaviors did *not* predict the degree to which these maternal behaviors actually occurred in everyday mother-child interactions. For some areas of maternal behavior, then, the verbally-expressed needs of women seem to be as good, or better, predictors of their actual maternal behavior practices than are these women's subjective assessments of their own childrearing practices.

Returning to the results of the present study, the fact that the Edwards PPS *n* autonomy and *n* dominance scales were not, as hypothesized, generally predictive of the *S*s' overt maternal behavior may have been due to a number of factors including: (a) poor construct validity of the scales for the population sampled, (b) poorly formulated hypotheses, and/or (c) a lack of "true relations" between the variables measured. Answers to these alternative possibilities await future research. At the moment, however, the findings of the present study warrant these general conclusions: (a) Middle-class women's verbally-expressed needs for aggression and nurturance clearly predict the degree that these women *overtly* express non-affectionate and nurturant behaviors toward their children. (b) There is a lesser trend that the *S*s' *n* autonomy is associated with nonnurturant aspects of their maternal behavior. (c) No evidence was found that the *S*s' stated desires to dominate others were directly translated into coercive maternal behavior practices.

SUMMARY

This study explored relations between certain verbally-expressed needs of middle-class women (as measured by their Edwards PPS responses) and the overt behavior these women displayed toward their children. Data regarding the latter were obtained from home visit observations of mother-child interactions and subsequent ratings of these maternal behaviors using the Fels Parent Behavior Rating Scales. Four general hypotheses were tested. These were: *Hypothesis I:* "Middle-class women's verbally-expressed *n* aggression will be negatively correlated with the amount of affection they display toward their children." *Hypothesis II:* "The *S*s' expressed need to nurture others will be positively associated with the protectiveness they exhibit with their children." *Hypothesis III:* "Women with high *n* autonomy will be less protective of their

children than will women with low *n* autonomy." *Hypothesis IV:* "Women who express strong needs to dominate others will be particularly prone to employ coercive maternal controls in their everyday child-socialization practices." Hypotheses I and II were clearly substantiated. Hypothesis III found limited substantiation; hypothesis IV, none. Implications of the findings were discussed.

REFERENCES

Baldwin, A., Kalhorn, J., & Breese, F. The appraisal of parent behavior. *Psychol. Monogr.*, 1949, *63,* No. 4 (Whole No. 299).

Crandall, V. J., & Preston, Anne. Patterns and levels of maternal behavior. *Child Develpm.*, 1955, *26,* 267–277.

Champney, H. The measurement of parent behavior. *Child Develpm.*, 1941, *12,* 131–166.

Edwards, A. L. *Manual; Edwards Personal Preference Schedule.* New York: Psychological Corp., 1954.

Lorr, M., & Jenkins, R. L. Three factors in parent behavior. *J. consult. Psychol.*, 1953, *17,* 306–308.

Roff, M. A factorial study of the Fels Parent Behavior Scales. *Child Develpm.*, 1949, *20,* 29–45.

30

Father-Absence and Delay of Gratification: Cross-Cultural Comparisons[1]

Walter Mischel

This paper by Mischel attempts to utilize a small segment of the child's behavior—the ability to delay gratification—in order to understand the effects of (1) the absence of the father and (2) certain class and ethnic factors. Although this study closely resembles an experiment, the difference between the groups is the social setting rather than any introduced variable. The ability to delay gratification has importance in psychoanalytic theory and also implications for education.

The dual purpose of this study is to investigate the effect of father-absence in the home on children's ability to delay gratification, as measured by their preference for immediate, smaller reinforcement (ImR) over delayed, larger reinforcement (DelR); and test hypotheses based on ethnographic data concerning cross-cultural and intracultural differences in preferences for ImR or DelR. Opportunities to study systematically the effect of father-absence on the child's behavior, in this case on his delaying the immediately available but relatively trivial for the sake of later but larger outcomes, are rare. Either fathers are not absent for long periods or other conditions tend to be so deviant that generalizations to less extreme cases must be most tentative. The possible effect of prolonged absence of a parent is nevertheless of such import for any theory of identification and personality development that data are much needed. . . .

The present research, based on investigations in the southern Caribbean, presents data on a large number of father-present and father-absent boys and girls, sampled from a variety of settings, and differing sufficiently in age so that the role of the latter can be tested. Father-presence or absence is related in this study to the individual's choice preference for ImR or DelR. Such choice preferences are fairly easily elicited and provide lifelike behavior which is readily quantified and some of whose empirical correlates have been explored independently. Namely, preference for DelR (as opposed to ImR) has been related positively to accuracy in time statements and to social responsibility (Mischel, 1961b). The latter was measured by an independently validated scale and found to correlate substantially with other measures of personal and social adjustment and maturity. Preference for DelR has also been negatively related to delinquency, positively to strength of n Achievement, and negatively to acquiescence or "yeasaying" (as opposed to "naysaying") tendencies (Mischel, 1961a). The preference for DelR-ImR distinction may thus be thought of as delineating two empirically-elaborated clusters, associated with significant differences in "maturity," responsibility, delay over time, long-term goal direction, and autonomy.

. . . Conceptually, the empirical correlates of preference for DelR and ImR are related to

---

From *Journal of Abnormal and Social Psychology,* 1961, 63, 116–124.

[1] This study was carried out under the partial support of a grant from the Laboratory of Social Relations, Harvard University, and Grant M-2557 (A) from the National Institute of Mental Health.

Freud's (1922) distinction between functioning on the pleasure principle as opposed to the reality principle and are relevant to Mowrer and Ullman's (1945) position. The latter maintain that the inability to delay gratification is an important factor in immature, neurotic, and criminal behavior.

The data on preference for DelR or ImR from the previous studies in this program all come from the same general culture area outside the United States and are based on West Indian children's responses, tested in comparable settings and with techniques similar to those used in the present study. Thus a network of correlates associated with the two reinforcement preference patterns with respect to delay has begun to be established. The findings of this study on choice behavior of father-absent and father-present children become meaningful in that context. In an earlier exploratory study (Mischel, 1958), using a small sample of Trinidadian rural children aged 7–9, a significant relationship was found between father-absence and preference for ImR. Further, Negro subjects were found to show greater ImR preference when compared with East Indian subjects in the same cultural setting, the differences between the two groups seemingly being due to greater father-absence in the Negro sample. The present study is in part directed at replicating this relationship and testing its generality by sampling male and female children from a large variety of settings.

More important, however, is the attempt to test also the effect of father-absence on older children. Thus far investigations on the effects of father-absence (Bach, 1946; Lynn & Sawrey, 1959; Sears, Pintler, & Sears, 1946; Tiller, 1958) have all been based on samples of children under age 10. Lynn and Sawrey (1959) summarized these findings. Their own research included hypotheses of greater immaturity and poorer peer adjustment on the part of father-absent boys as compared to father-absent girls and a control group of father-present boys. These (and other) hypotheses were generally supported by data based on Norwegian boys and girls, ranging in age from 8 years to 9 years 6 months. The present study offers data for this age group (ages 8 and 9) as well as for a second, older group, aged 11–14. Lynn and Sawrey rightly point out that the findings of the earlier studies are limited by the fact that the father-absence with which they were concerned was a temporary state of affairs during wartime. In contrast, Lynn and Sawrey's subjects came from Norwegian sailor families where the father is often away from home for periods of 2 years or more with only infrequent visits. These authors recognize that "unknown cultural factors specific to the samples may be operating instead of or in addition to the variable of father-absence. Women who marry sailors may, for example, be such a select group that the variable 'sailor wife' may be the crucial one" (p. 259). The present data on father-absence from the Caribbean have the additional advantage of coming from a culture in which father-absence is a common and widely accepted pattern. Common-law marriage is extremely popular and according to some estimates a permanent father or father-figure, is prolongedly absent in over one-third of all households.

The general hypothesis here expects an inverse relationship between father-absence and preference for DelR. Father-absent children should show relatively less preference for DelR and relatively more preference for ImR in choice situations, when compared with comparable father-present children. This relationship is expected to reflect the relatively greater immaturity and other correlates of preference for ImR which, it is anticipated, should characterize the father-absent group more than the father-present group.

The second aim of the study is to test experimentally an anthropological hypothesis concerning gross personality (or "national character") differences with respect to preference for DelR or ImR existing between the populations of two islands, Trinidad and Grenada, both in the southern Caribbean. The two islands may be thought of as subcultures within the same larger culture of the West Indies. The significance of this second aim lies in the application of quantitative, experimental techniques for testing a cross-cultural hypothesis, in this case a hypothesis dealing with an important psychological dimension that has been used frequently to make global but largely untested characterizations of groups and societies. This requires some explication of the cultural contexts within which the anthropological hypothesis was formulated.

On the basis of anthropological observation, a major personality difference between the Negro groups of the two islands was noted. This difference, as expressed by many informants with experience in both subcultures, is that the Trinidadian tends to be more impulsive, indulges himself, and settles for relatively little if he can get it right away. He is described as not working or waiting for larger rewards in the future, but tending instead to prefer immediately available relatively trivial gratifications. In contrast to this, the Grenadian, although on the whole poorer, is said to be more willing and able to postpone immediate gains and gratifications for the sake of larger rewards and returns in the future.[2] These observed personality or "national character" differences are supported by the further observations that the Trinidadian's savings (e.g., for the education of children) tend to be less than the Grenadian's of equally poor or even poorer circumstances.

Such differences between the islands may be related historically to the fact that the Grenadian, in contrast to the Trinidadian, has tended to hold on to his own land from generation to generation, and has been able to develop a much more autonomous role with respect to dependence on non-Negroes and colonials in government, in the professions, and in all forms of business.

The observed distinction between Trinidadians and Grenadians may be thought of as reflecting differences with respect to preference for ImR as opposed to DelR in choice situations, with the Trinidadian preferring the former more and the latter less than the Grenadian. This difference between individuals from the two islands may be conceptualized, at least in part, as related to differences in "trust" experiences with respect to the actual anticipated occurrence of delayed reward.

Indeed, in accord with Rotter's (1954) social learning theory we have conceptualized earlier (Mischel, 1958) that the person's choice of an immediate, smaller or delayed, larger reward in a choice situation is, at least to some extent, a function of his expectation that the reward will actually occur and the reward value, or preference value, of the particular reinforcement. Mahrer (1956) has already shown that the person's expectancy that reinforcement would follow from the social agent making the promise, this kind of behavior may be thought of as "trust" or the belief that the agent promising delayed reinforcement will actually supply it. even after time delay, is an important variable in such choice behavior. In common sense terms,

The Grenadian may be thought of as having had a longer history of actual reward occurrence in delay situations (e.g., in the form of long-term gratifications accruing from land ownership) and as living in a cultural situation in which there is a good deal of trust with respect to promise-keeping and with respect to other means through which reward delay may be mediated (i.e., long-term "payoff"). Certainly direct observation suggests the Grenadian to be far less suspicious and skeptical in his relations to strangers as well as to his own peers than is the Trinidadian. In sum, the former may be thought of as having long experiences in a relatively stable "delayed reward culture" (at least in its economic aspects) within which he himself has had a relatively autonomous and "trusted" role; in contrast, the Trinidadian may be thought of as participating in a relatively "immediate reward culture" within which he has had a highly dependent and "untrusted" role. These differences should, in turn, be reflected in the expectations with respect to reward delay transmitted to the children within the two cultures. The prediction follows that there would be greater preference for ImR than DelR on the part of Trinidadian as compared to Grenadian children of comparable age and socioeconomic background.

Lastly, a hypothesis concerning differences between the Negro and East Indian populations of the same island, Trinidad, on the same dimension of preference for ImR or DelR is tested. Ethnographic observation suggested differences between these two intra-island groups similar to those noted between Trinidadians and Grenadians (where reference is made only to the Negro populations of both islands). Namely, when compared to Trinidadian Negroes, the East Indian group of Trinidad appears to show clearly greater preference for DelR in daily behavior. As previously indicated, this observation has been made and supported in earlier research (Mischel, 1958). Here an attempt is made to replicate these findings, taking into account any possible differences in extent of father-absence in the samples tested.[3] This also permits comparisons of intra-island differences between large groups on the dimension of concern.

---

[2] The distinction in this respect between Trinidadians and Grenadians has been made previously to differentiate the Negro and East Indian subgroups within Trinidad (Mischel, 1958). and will be referred to again, with replicating data, later in this paper.

[3] With respect to household structure in the three samples studied, the following tentative observations were made. Trinidad and Grenada are alike in the household structure of lower class and lower middle class Negro families. In both islands common-law marriage is extensive and accepted, fathers are frequently prolongedly or permanently absent, and the mother-child relationship is the family nucleus. The child usually knows who his natural father is or was, although the mother's partner may change at irregular intervals. In contrast, in East Indian Trinidadian families the father tends to be permanent and common-law marriage is rare.

METHOD

Subjects

A total of 68 Trinidadian Negro children (30 boys and 38 girls), and of 69 Grenadian Negro children (36 boys and 33 girls), all in the age group 8 through 9 were tested. The Trinidadian subjects came from two schools, both on the outskirts of the capital city. The Grenadian subjects similarly came from two schools, both located on the outskirts of Grenada's capital city. Care was taken to select all schools so that subjects of roughly comparable socioeconomic backgrounds (all lower middle and lower class) would be obtained. This selection was guided by impressions formed on the basis of inspection of a variety of schools, participant observation in both cultures, and the counsel of local school officials. Inspection of the father's occupation reported by father-present subjects in all samples reveals no striking differences. Over 90% of the responses seemed classifiable as "unskilled workers."

Further data were collected on two samples of Trinidadian East Indian subjects aged 8 through 9 from the same Trinidadian schools. The first sample consisted of 22 boys and 15 girls; the second of 20 boys and 23 girls. In addition to these young subjects, similar data were collected from a total of 112 older Trinidadian Negro subjects (68 boys and 44 girls) aged 11 through 14 at one of the Trinidadian schools indicated above, and on 75 Negro subjects (40 boys and 35 girls) in the same age range, coming from the second Trinidadian school.

Task and Measures

The reinforcements used with the younger group were selected through pretesting with a random sample of 20 boys and girls, aged 8 and 9, tested at another government school on the outskirts of the capital city in Trinidad. In these pre-experimental sessions subjects were seen in individual sessions and their preferences for specific reinforcements, all candy, were elicited. As a result, two candy reinforcements, varying markedly in price, size, and packaging (i.e., a \$.02 and a \$.10 candy bar) were selected. These met the desired requirements inasmuch as the larger reinforcement was uniformly preferred in a straight choice situation ("which *one* of these two would you like to take?"), but when the choice was "you can have this one (the smaller) today *or* this one (the larger) in one week," approximately 50% of the group chose the former and approximately 50% the latter. The purpose of this procedure was to select a reinforcement pair which would, as closely as possible, dichotomize the choices of subjects in one of the two ethnic groups being compared supplying an approximate 50-50 split in the Trinidadian sample which could be used as the criterion for comparison with the Grenadian data.

With the older subjects, a different reinforcement pair, namely, a \$.10 and a \$.25 candy bar were used on the basis of the procedure reported earlier (Mischel, 1961b). In addition, two verbal items, to which the subject was to respond with "yes" or "no," one inserted near the beginning of the total procedure and the other in the middle, were included with the older subjects. The items were:

1. I would rather get ten dollars right now than have to wait a whole month and get thirty dollars then.
2. I would rather wait to get a much larger gift much later than get a smaller one now.

The behavioral choice (candy) of preference for ImR as opposed to DelR was given as the last item of the total battery. Previous research (Mischel, 1961b) has shown that the use of these three measures in combination is more fruitful than the use of single items.

Although there are no formal data available on the comparative subjective value of the candy bars in the two cultures to be compared, direct observation, confirmed by those with intimate knowledge of both cultures, indicated that in both Grenada and Trinidad the particular candies used were of similar appeal and availability. In both cultures such candies are sold in virtually every corner store and are popular with, and commonly consumed by, middle and lower class children alike. Candy allowances in the two cultures appear to be essentially the same.

The measure of father-absence or presence consisted of the response ("yes" or "no") to the question: "Does your father live at home with you?" Twenty-eight out of 30 such responses, randomly selected, were in agreement with the child's mother's response to the question, "Does his (specifying the child's name) father live at home with you?" asked in subsequent interviews with mothers of subjects in the older Trinidadian sample.

Procedure

The same experimenter was used in all testing. He was introduced as an American from a college in the United States interested in gathering information on the children in the various schools of the island. To help with this, subjects were asked to answer a number of simple questions, e.g., name, age. The details of this procedure have been described previously (Mischel, 1958).

Upon completion of the questions, the experimenter expressed his wish to thank the group for their cooperation. He displayed the two kinds of reinforcement and said: "I would like to give each of you a piece of candy, but I don't have enough of these (indicating the larger, more preferred reinforcement) with me today, So you can either get this one (indicating the smaller, less preferred reinforcement) right now, today, or, if you want to, you can wait for this one (indicating) which I will bring back next Friday (one week delay interval)." These instructions were repeated in rephrased form to insure clarity and both reinforcements were carefully displayed. It was stressed that getting the (smaller) candy today precluded getting the (larger) one next week and vice versa. Subjects were asked to indicate their choice by writing T (today) or F (Friday, next week) on their questionnaires. The response made here was the measure of preference for ImR or DelR. Subjects were seated sufficiently far apart from each other to insure reasonably that their choices were made independently. All testing was done in group settings, and conducted on the same weekday (but in different weeks) and at similar times of day, within each of the schools in large classrooms.

RESULTS AND DISCUSSION

The data from the present study were analyzed in terms of differential preference for ImR as opposed to DelR in relationship to cultural group (Trinidad as compared to Grenada), and in relationship to presence or absence of the father within the home in each culture. In addition, the relationship between the experimental choice of preference for ImR as opposed to DelR and presence or absence of the father in the home was examined in data from the group of older Trinidadian subjects, for whom a Grenadian comparison group is not available. All statistical analyses were made with the chi square test.

Careful examination of possible sex differences with respect to either extent of father-absence or preference on the experimental choice, within each of the subsamples and in combination, revealed no differences approaching statistical significance. Indeed, no consistent sex trends were observed in examining all reported relationships separately for each sex. Consequently, the male and female data were combined for all subsequent statistical analyses. The similar choice preferences with respect to delay of father-absent boys and girls in this Caribbean sample is of interest in view of Lynn and Sawrey's (1959) Norwegian data, in which differential effects of father-absence on the sexes were found.

The data used for relating the experimental choice with cultural group are given in Table 1, which shows the number of Negro subjects (ages 8 through 9) within each of the schools, in Trinidad as compared to in Grenada, preferring ImR or DelR. Comparison of the number of subjects preferring ImR or DelR *within* each of the two cultures (i.e., comparing School 1 with School 2 in Grenada and School 3 with School 4 in Trinidad) results in no chi square which reaches significant probability levels, and consequently the data from the two schools within each culture can be combined for further comparisons. This combination, namely, a comparison of the total number of Negro Trinidadian subjects preferring ImR or DelR with the total number of Negro Grenadian subjects preferring ImR or DelR, results in a chi square (using Yates' correction for continuity) of 10.40, with $p < .01$. As predicted by the cultural hypothesis, a significantly larger proportion of the Trinidadian subjects, as compared to the Grenadian subjects, chose ImR rather than DelR.

The data used for relating father-absence and the experimental choice, in Trinidad and in Grenada, are also presented in Table 1. The chi square test applied to the relevant data yields uncorrected chi square values of 6.14 for Trinidad and of 5.15 for Grenada. These values are significant ($p < .02$ and $p < .05$, respectively) and indicate a greater proportion of father-absent subjects, as compared to father-present subjects choose ImR rather than DelR in each culture. If extreme caution is used and the Yates correction is employed on all fourfold contingency tables (in spite of the fact that all expected and marginal frequencies are relatively large) the chi square value for Trinidad becomes 4.93 ($p < .05$) and for Grenada 3.80. A value of 3.84 is required for $p < .05$ when $df = 1$.

In view of this support for the relationship

Table 1: *Number of father-present (Fa +) and father-absent (Fa −) subjects choosing immediate, smaller or delayed, larger reinforcement*[a]

Reinforcement Preference	GRENADA (NEGRO)						TRINIDAD (NEGRO)					
	School 1		School 2		Total[b]		School 3		School 4		Total[c]	
	Fa +	Fa −	Fa +	Fa −	Fa +	Fa −	Fa +	Fa −	Fa +	Fa −	Fa +	Fa −
Immediate, smaller	5	4	4	4	9	8	11	10	8	7	19	17
Delayed, larger	16	6	26	4	42	10	16	6	10	0	26	6

Note.—Grenada versus Trinidad × Reinforcement Preference: $\chi^2 = 10.40$, $p < .01$ (corrected for continuity).

[a] Subjects are aged 8 and 9.

[b] $\chi^2 = 5.15$, $.05 > p > .02$; $\chi^2 = 3.80$, $.10 > p > .05$ if corrected for continuity.

[c] $\chi^2 = 4.93$, $.05 > p > .02$ (corrected for continuity).

between father-absence and preference for ImR, it becomes important to examine the differences between the two cultures with respect to degree of father-absence. Table 1 shows the number of father-absent and father-present subjects within each of the samples from the two cultures. Among the Grenadians, 18 of the 69 subjects reported father-absence; among the Trinidadians, 23 of the 68 subjects reported father-absence. Application of the chi square test to these data results in no p values approaching statistical significance, and suggests that the difference between the two cultures on the experimental choice cannot be attributed to differential father-absence although within each culture father-absence is related to the choice behavior.

This finding is of special interest in view of the finding obtained earlier (Mischel, 1958) in a comparison of Negro and East Indian subjects, both within the Trinidadian culture, on preference for delayed reinforcement. Although the two Trinidadian subgroups differed significantly in the choice behavior (the Negro subjects more frequently preferring the immediate, smaller reinforcement), the differences appeared attributable to differences in father-absence (the Indian subjects being characterized by more father-present households). The present data permit an attempt to replicate this relationship. The new data for Trinidadian East Indian subjects aged 8 through 9 were as follows. Among the Indians, 26 chose the ImR and 54 the DelR. Only 4 subjects in the Indian group were father-absent, 2 of these choosing ImR and 2 DelR. Comparing the Indians with the Trinidadian Negro subjects on preference for ImR as opposed to DelR results in a corrected chi square of 5.50 ($p < .02$). As was found earlier, the East Indian group again shows greater preference for DelR as opposed to ImR relative to the preferences of the Trinidadian Negro group. Note that there are no significant differences in reinforcement preference with respect to delay between the Trinidadian Indian group and the Grenadian group, although the latter shows significantly greater father-absence (chi square = 11.47, $df = 1$, $p < .001$). Comparison of the Trinidadian Indian group with the Trinidadian Negro group on extent of father-absence results in a corrected chi square of 18.5 ($df = 1$, $p < .01$); clearly the latter show significantly greater father-absence than the former. If *only* father-present subjects from the two groups (Trinidadian Indian and Trinidadian Negro) are compared on preference for ImR or DelR, no significant relationship holds. In contrast, comparison of only father-present Trinidadian Negros with Grenadian father-present subjects on reinforcement preference still yields a corrected chi square of 5.85 ($p < .02$), the Grenadians showing greater DelR preference. Thus the intra-island observed differences between the two Trinidadian groups appear to be largely related to differential father-absence between the two groups. However, the Trinidadian-Grenadian Negro differences with respect to reinforcement preference cannot be attributed to differential father-absence.

Summarizing the results thus far, the following relationships, condensed in Table 2, should be noted. Grenadian Negro children showed significantly greater preference for DelR as opposed to ImR when compared with Trinidadian Negro subjects. This difference cannot be attributed to differential father-absence since the two groups show no significant difference with respect to that variable, and since the difference in DelR-ImR preference holds even when only father-present subjects from the two islands are compared. In contrast, comparisons between Trinidadian Negro subjects and Trinidadian East

Indian subjects reveal that the difference between these two, namely, the relatively greater DelR preference on the part of the East Indian group, does appear attributable to differential father-absence. This seems the case since the Trinidadian Negro subjects show significantly greater father-absence than do the East Indian subjects, and since the latter's relatively greater DelR preference does not hold when this differential father-absence is taken into account, i.e., when comparisons are based only on father-present subjects. Also note that Negro subjects from Trinidad and Grenada do not differ in their choice behavior *if* only father-*absent* Negro subjects from the two islands are compared (see data in Table 1). The small *N* of father-absent East Indian subjects prevents their being compared with other father-absent groups. Finally, the Trinidadian East Indian group and the Grenadian Negro group are not significantly different in their reinforcement preferences, although the latter show significantly greater father-absence.

The present results thus suggest that the difference between the Trinidadian and Grenadian subjects on the experimental choice must be sought in sources other than family constellation with respect to father-absence or father-presence. This illustrates a situation commonly encountered in cross-cultural research. In spite of overlapping rearing conditions, not merely in terms of the father in the family constellation but also in terms of numerous patterns that link the two cultures and make them homogeneous in many ways, other unspecified sources of variance must account for gross observed differences, in this case differences in the delay of gratification patterns. The writer is inclined to focus on some of the cultural and economic differences, especially differences in autonomy and independence, as indicated earlier. The lack of a significant difference between father-*absent* Trinidadian and Grenadian Negro subjects in preference for DelR as opposed to ImR (Table 2) also suggests the possible argument that when fathers are absent differences in cultural values are less likely to be internalized by the children, since fathers may be needed to transmit such cultural values.

We turn next to the data on the relationship between father-absence and reinforcement preference obtained from the 112 older (ages 11 through 14) Trinidadian subjects. Here, as previously described, three measures of reinforcement preference with respect to delay were available. Application of the chi square test resulted in no p value less than .30. This lack of relationship was obtained in comparisons using the candy choice alone, as well as all possible combinations of preferences over the three measures. To inquire further into this failure of the anticipated relationship for older subjects, the same experimental procedure was repeated on another sample of 75 Trinidadian subjects of the same age range, tested in another but highly similar Government school, comparably located, and with subjects of similar socioeconomic and ethnic backgrounds. Again, no relationship between the experimental choices and presence or absence of the father in the home was obtained that reached or approached significant probability

Table 2: *Summary of relationships between ethnic group,*[a] *subjects with father-present* (*Fa* +) *or father-absent* (*Fa* −), *and preference for delayed, larger* (*DelR*) *as opposed to immediate, smaller* (*ImR*) *rewards*

Comparison	Greater DelR Preference	p value[b]
Trinidadian Negro vs. Grenadian Negro (Fa + and Fa−)	Grenadian	< .01
Trinidadian Negro vs. Grenadian Negro (Fa + only)	Grenadian	< .02
Trinidadian Negro vs. Grenadian Negro (Fa − only)	—	*ns*
Trinidadian Negro Fa + vs. Fa −	Fa +	< .05
Grenadian Negro Fa + vs. Fa −	Fa +	< .07
Trinidadian East Indian vs. Trinidadian Negro (Fa + and Fa −)	East Indian	< .02
Trinidadian East Indian vs. Trinidadian Negro (Fa + only)	—	*ns*
Trinidadian East Indian vs. Grenadian Negro (Fa + and Fa −)	—	*ns*
Trinidadian East Indian vs. Grenadian Negro (Fa + only)	—	*ns*

Note.—Comparisons with East Indian subjects on father-absence in relation to other variables cannot be made in view of the small N (4) of such father-absent subjects. Trinidadian Negro subjects as well as Grenadian Negroes, show significantly greater father-absence ($p < .01$) than Trinidadian East Indian subjects, but the difference between the two Negro groups in extent of father-absence is ns.

[a] *Using Trinidadian Negro, Trinidadian East Indian, and Grenadian Negro subjects, all aged 8 and 9.*

[b] *Based on chi-square comparisons, corrected for continuity,* ns = *p value > .10.*

levels. Thus although the expected relationship was repeatedly obtained at the younger age levels, the null hypothesis cannot be rejected for seemingly similar subjects at older age levels.

The possibility exists that the instructions, for the experimental choice, specifically the phrase "but I don't have enough of these (DelR) with me today," may have partially confounded the results by producing a conflict in older subjects, eager for the DelR yet reluctant to make the experimenter return. However, this seems extremely unlikely. First, during the administration, that phrase, used only as an introduction to make the choice plausible, was probably overshadowed by the frequent repetitions of the actual choice conflict and the implication that the experimenter would *definitely* return next week to hand out the DelR to those in the group who chose it. Second, and more definitive, in over 20 pre-experimental sessions with other comparable subjects from the same culture, used to select the particular reinforcements sets, post-choice inquiries into reasons for the choices yielded nothing remotely suggestive of such an "obliging response."

It is tempting to argue that for older subjects the measures of preference for ImR as opposed to DelR is meaningless or invalid, and consequently relations with father-absence should not be expected. It must be recalled, however, that the same measures used with Caribbean subjects of this age, and indeed including subjects from the present older sample, have yielded significant correlations with conceptually relevant variables (social responsibility, need for achievement, non-acquiescence), all of the above being positively correlated with preference for DelR (Mischel, 1961a, 1961b). In this context, it should also be noted that no relationships approaching statistical significance were obtained in attempts to relate father-absence with these correlates of preference for DelR. This suggests that father-absence at older age levels may not have some of the effects (e.g., "immaturity," poorer adjustment) apparently associated with it at the earlier age levels (e.g., Lynn & Sawrey, 1959). The question cannot be answered at present in view of the lack of research on father-absence with older children. Of course, the present data are seriously hampered in the gross measure of father-absence or presence used, which gives no information on the extent or duration of the father's absence or presence. Nevertheless, the same gross measure did yield the expected results with the younger subjects, and excellent agreement between older subjects' reports of father-absence and mother's reports on the same question was obtained.

The *post hoc* interpretation being suggested for these findings is that as the individual develops and matures, he begins to participate in an environment that extends beyond the immediate family and consequently his expectations with respect to promise-keeping, his "trust," and his choice behavior with respect to delay of gratification become contingent upon numerous factors and experiences other than those within the household itself. Similarly, at least in southern Caribbean cultures, sources of influence outside the immediate family may become important determinants of other variables correlated with ImR-DelR preferences, such as "social responsibility" and need for achievement, at relatively early periods in the person's life.[4]

Although this interpretation is clearly tentative, pending further and more controlled research, the data illustrate the complexity of some of the conditions that may influence such choice behavior. In spite of these complexities, significant support was obtained for the anthropological hypothesis concerning cross-cultural differences between Trinidad and Grenada with respect to delay of gratification. Earlier findings of differences with respect to ethnic groups on the same variable were replicated. It is recognized, of course, that the findings are necessarily limited by sampling problems which make any generalizations tentative. Nevertheless, the findings are construed as illustrative of the possibilities of applying quantitative simple experimental techniques for testing cross-cultural differences on important psychological dimensions.

SUMMARY

This study tested an anthropological hypothesis concerning differences in preference for immediate, smaller (ImR) as opposed to delayed, larger reinforcement (DelR) between two subcultures (Trinidad and Grenada) within the West Indies. The data, in the form of choice

---

[4] It should be noted that in the cultures studied, the individual becomes independent of the family and autonomous, e.g., vocationally, relatively early in life. For example, official "student-teachers" in public schools are often 13 years old. In all three samples, the child from age 10 or 11 on, becomes increasingly free to move about, visiting with other relatives (e.g., uncles, grandparents, etc.) who may or may not be living in the immediate neighborhood, and whom he may visit for a day or longer periods.

preferences for ImR or DelR elicited from children in both islands, supported the cultural hypothesis. Further, a significant relationship between absence of the father within the home and greater preference for ImR as compared to DelR was found in both cultures, using subjects aged 8-9. However, this relationship was not found for older subjects (aged 11-14). These results were discussed in terms of increased sources of variance relevant to the formation of trust behavior and expectancies outside the household with increasing age. Data replicating previously obtained differences with respect to preference for ImR as opposed to DelR between two ethnic groups (Indian and Negro) within the island of Trinidad were also presented.

REFERENCES

BACH, G. R. Father-fantasies and father typing in father-separated children. *Child Develpm.*, 1946, *17*, 63–80.

FREUD, S. *Beyond the pleasure principle.* New York: Boni & Liveright, 1922.

LYNN, D. B., & SAWREY, W. L. The effects of father-absence on Norwegian boys and girls. *J. abnorm. soc. Psychol.*, 1959, *59*, 258–262.

MAHRER, A. R. The role of expectancy in delayed reinforcement. *J. exp. Psychol.*, 1956, *52:* 101–105.

MISCHEL, W. Preference for delayed reinforcement: An experimental study of a cultural observation. *J. abnorm. soc. Psychol.*, 1958, *56*, 57–61.

MISCHEL, W. Delay of gratification, need for achievement, and acquiescence in another culture. *J. abnorm. soc. Psychol.*, 1961, *62*, 543–552. (a)

MISCHEL, W. Preference for delayed reinforcement and social responsibility. *J. abnorm. soc. Psychol.*, 1961, *62*, 1–7. (b)

MOWRER, O. H., & ULLMAN, A. D. Time as a determinant in integrative learning. *Psychol. Rev.*, 1945, *52*, 61–90.

ROTTER, J. B. *Social learning and clinical psychology.* New York: Prentice-Hall, 1954.

SEARS, R. R., PINTLER, MARGARET H., & SEARS, PAULINE S. Effect of father separation on preschool children's doll play aggression. *Child Develpm.*, 1946, *17*, 219–243.

TILLER, P. O. Father absence and personality development of children in sailor families. *Nord. psykol. Monogr.*, 1958, Series No. 9.

THE CHILD IN SCHOOL

31

The Comparative Influence of Punitive and Nonpunitive Teachers upon Children's Concepts of School Misconduct[1]

Jacob S. Kounin and Paul V. Gump

It has long been understood that the behavior of the teacher influences the behavior of the child; but—as in the case of knowing that parent behavior influences the child—the particulars have not been clearly understood. This study by Kounin and Gump is one of a series attempting to analyze, in greater detail, just what particular behaviors of teachers influence what specific concepts children hold about school. In this attempt the authors investigated the attitudes of first-graders toward misconduct based upon the experiences these children were having with punitive and nonpunitive teachers.

From *Journal of Educational Psychology*. 1961, *52:* 44-49.

[1] A version of this paper was presented at the American Psychological Association meeting, September 5, 1959. It is part of an investigation supported in part by a Research Grant, M–1066, from the National Institute of Mental Health, United States, Public Health Service and in part by a grant from the College of Education of Wayne State University.

This paper reports a portion of a research project pertaining to the management of childrens' behavior in classroom settings. Because so many teachers, especially beginners, verbalize considerable concern about discipline and control, we are focusing our current research in this area. While there is some relevant literature, based upon experience and insightfulness, we have been unable to locate any generalizations based upon data from research.

In a previous study by Kounin and Gump (1958) specimen-record types of observations were gathered of discipline incidents during the first week of kindergarten, focusing upon the triad of: a misbehaving child (target), a teacher doing something to stop the misbehavior, and a watching audience-child. Limiting our dependent variables to overt behavior we found that teachers' techniques of handling a misbehaving kindergarten child (target) did have different degrees of socializing success upon audience-children. A socializing success was defined as an observable reduction of overt misbehavior or an increase in conforming behavior (standing up "even straighter" in line). Control techniques high in clarity (defining the deviancy, specifying how to stop) were most successful. Control techniques high in firmness (standing closer to the misbehaving child, continuing to look at him until he stopped misbehaving) were successful only for audience-children who were themselves deviancy-oriented at the time. Control techniques high in roughness (anger, physical handling) were least successful and tended to be followed by behavior disruption (less involvement in work, overt signs of anxiety) rather than conformity on the part of audience-children. In terms of their effects, it is evident that roughness is a different dimension than firmness. . . .

It is postulated that aggression leads to counteragression; it is further postulated that a punitive teacher has more power over her pupils than they have over her and that she blocks overt manifestation of pupils' aggression (observations in the classrooms of the punitive teachers selected for this study indicate that this second assumption is tenable). From these two postulations, we derive the following hypotheses:

1. That the school misconduct preoccupations of children with punitive teachers will contain more aggression than those of children with nonpunitive teachers.
2. That children with punitive teachers will be more conflicted about school misconduct than will children with nonpunitive teachers.
3. That the aggression needs and the conflict relating to misconduct hypothesized to exist among children with punitive teachers will detract from their concern with school-unique values that are not directly related to misconduct.
4. The question may also be raised as to whether or not the amount of tension generated in the children with these particular punitive teachers is sufficiently great to reduce the rational qualities of their attitudes toward misconduct.

METHOD

Subjects

The subjects were 74 boys and 100 girls attending their first semester of the first grade in the public schools of a large city. They represented all the children from six home rooms of three schools, in from upper-lower to middle-middle socioeconomic neighborhoods.

Procedure

Overall school climate was controlled by selecting pairs of punitive vs. nonpunitive teachers from the same school. Three such pairs were obtained from three elementary schools.

The initial selection of punitive and nonpunitive teachers was obtained from the principal and assistant principal. Following this the classes were observed by both principal investigators. At approximately a week later the teachers were further rated by a supervisor of student teachers who visited each class twice.

The raters checked along a continuum from Extremely Punitive (threatens children with consequences that really hurt; makes threats that imply sharp dislike, real willingness to harm child; ever-readiness to punish) to Not Punitive (does not punish and does not threaten). A punitive vs. nonpunitive pair of teachers was used for the study only when all five persons agreed on their dichotomizations. All the teachers were rated as having good organization, well-behaved classes, and as achieving the learning objectives for their grade. Eighty-four of the children were in classes with punitive teachers and 90 children were in classes with nonpunitive teachers.

The children were interviewed individually during the third month of attendance at school. The interview consisted of the questions: "What is the worst thing a child can do at school?"

and, following the reply, "Why is that so bad?" Identical questions were asked regarding home as the milieu for misbehavior.

Coding the Replies

The misconducts mentioned by the children were coded for content and for certain qualities or dimensions.

The content code (obtained from the question of "What is the worst thing to do?") contained two parts: the misconducts and the explanations given for why these were bad. The misconduct included: the act type (physical or psychological assaults, noncompliance, etc.) and the object of the misconduct (parents or teachers, other children, institutional laws or custom, etc.).

The code for the explanation of misconduct was designed to answer three questions: Who is involved in the consequence (the child himself, parent or teacher, a peer, etc.)? What kind of sufferings result to others from the misconduct (physical pain, achievement loss, property loss, etc.)? What kinds of retributions occur to the misbehaver (work imposal, character loss, physical punishment, etc.)?

RESULTS AND CONCLUSIONS

Children probably answer the question of "What's the worst thing a child can do in school?" with a report of acts that reflect their preoccupations. It is not likely that our subjects' answers would have been the same if they were presented with a forced-choice of alternative acts. Given a choice, most children would probably rate "stabbing someone" as more serious than "talking in class." If the misconducts the children talked about are taken to represent tension systems and preoccupations, we may infer from these the comparative impact of punitive and nonpunitive teachers.

In a concurrent study of childrens' attitudes toward misconduct (201 boys and 214 girls in the first grade of six public schools representing a range of socioeconomic backgrounds), Gump and Kounin (1959) found both sex differences and differences between home and school milieus. For example, home misconducts included more breaking of objects while school misconducts included more rule violations; parents suffered more than teachers in consequences but teachers retributed more frequently. However, parents were reported as retributing with more corporal punishment and with more severe punishment than teachers. There were also differences in the responses of boys and girls, especially in school. For example, girls reported "talking" as a school misconduct eight times more frequently than boys, whereas boys reported physical assaults on peers in school more frequently than did girls.

Consequently, the comparison of the responses of children with punitive and nonpunitive teachers was made separately for sexes and also for home and school milieus. However, on *all* comparisons of school responses the direction of differences between children with punitive and nonpunitive teachers was the same for both boys and girls. The report of results, therefore, combines both boys and girls. Insofar as the differences between children with punitive and nonpunitive teachers are concerned, only 2 of the 48 comparisons of home responses were statistically significant: home misconducts of the children with punitive teachers were rated as more serious ($p < .05$, for girls only) and retributions to the subject were more serious ($p < .02$, for boys only). It is uncertain whether these represent some spillover of the influence of punitive teachers onto attitudes toward home misconducts, or whether they are chance differences for the number of comparisons made.

The results to be reported here, then, refer to boys and girls combined and to school misconducts only. Intercoder reliabilities ranged from 73-95% agreement, with a median of 90. The p levels of differences are based on the x^2 test. . . .

Following are definitions of the codes used which are not self-explanatory both for the misconducts and for the subjects' explanations for the wrongness of the act. These codes appear in Table 1.

I.

A. Physical assaults include all physical attacks on other persons (pushing, hitting).

B. Milieu-seriousness refers to the length to which the milieu would go to prevent such an act. The school would practically ignore "scratching head," would mildly frown at "whispering," and would go to any length to stop burning down buildings.

C. Coder seriousness refers to the general immorality or danger in the misconduct considered from the point of view of the overall culture. (The only frame of reference which produced high intercoder agreement was when the coders took the position of an understanding Reformed rabbi or a Unitarian minister. Taking an unspecified role, or that of either a parent or a teacher produced low intercoder agreement.) "Studying spelling lessons at the wrong time" is morally trivial while "maiming someone" is morally very serious.

D. Abstractness of misconduct refers to the size of coverage. It may range from a unique, "one time" misbehavior, such as, "cut a climbing rope in gym," to an abstract one, such as, "be mean to other people."

II.

A. A central adult is the responsible leader: teacher at school.

B. A psychological loss to another is exemplified by "It would make her worry."

C. Seriousness of consequences to others range from trivial harm, such as, "She'd be annoyed" to serious ones, such as, "He'd die."

D. A reality-centered retribution (this is scored only when the perpetrator himself suffers in the consequence) is coded when the consequence of a misconduct follows naturally from the act-type, such as: "not study because you'll get behind in your work." This contrasts with the response in which the connection between act and consequence is dependent upon a personal intervention of another, such as: "not study because teacher will make you stand in the corner."

E. "Reflexive justification" was coded when the child gave no consequence for either himself or others in his explanation of why the act was bad. When he said the act is bad because "It's not nice" or "It's bad" it was called a reflexive justification.

III.

A. On ego-acceptability, we sought to determine the degree to which the respondent could see himself as the perpetrator of the misconduct. In an ego-alien act, the respondent expresses abhorrence, such as: "It's dirty to hit little kids who didn't do nothing to you." An ego-attractive act is one in which the child indicates its seductive quality for him, such as: "Tell off a teacher —boy, I'd like to do that."

B. On the premeditation category, we sought to learn the extent to which the child sought to do wrong. If premeditated, the child plans the act and intends the consequences ahead of time, such as: "Put thumb tacks on teacher's chair when she is out." If intentional, the child accepts his part in the wrongdoing but does not plan it, such as, "talk during a lesson."

IV.

Aggression ("blood and guts") refers to the amount of aggression the respondent expresses in his misconducts and consequences. "Play in the storage bin because somebody might get hurt" expresses less aggression than "Play in the storage bin because you might push a kid off and there could be a sharp rock down there and he could hit his head against it and crash open his skull and he would bleed and his brains would fall out and he'd die."

V.

A. A learning or achievement loss is coded when interference with learning is the misconduct or the explanation, such as: "It's bad to make noise because somebody could make a mistake in his work," or ". . . because then he couldn't read good."

B. An institutional law violation is a violation of the rules of the school, such as: "talk when you're supposed to study," "not take your seat when the bell rings."

The results presented in Table 1 may be summarized around the three hypotheses and the one question raised in the introduction:

Punitive teachers will create or activate more aggression-tension than will nonpunitive teachers. This is strongly supported by the data. The children who have punitive teachers have more sheer aggression in their sins and consequences, they give both more milieu-serious and more coder-serious misconducts, their targets suffer more harm, they give more physical assaults as act-types, and their targets suffer more physical harm. The targets of children with nonpunitive teachers are more inclined to suffer psychological losses as consequences. As an example of the results: of 84 respondents with punitive teachers, 31 give physical assaults on other children and 40 mention school rule violations; while of 90 children with nonpunitive teachers, 15 talk about physical assaults and 56 about rule violations. (The remainder of the act-types are nonconformances and "miscellaneous.")

Children with punitive teachers will be more unsettled and conflicted about misbehavior in school. This hypothesis is supported by the findings related to the role of self in misbehavior. The children from nonpunitive teachers give misconducts in which their own role is intentional whereas children from punitive teachers give both premeditated and ego-alien misconducts. We may say that children with punitive teachers express more abhorrence for the misdeeds which they have selected and yet select misdeeds which require "malice and forethought."

Punitiveness of teachers will detract from children's concern with school-unique values. This hypothesis is supported. Children from punitive teachers talk more about physical attacks on peers—misbehavior by no means unique to the classroom setting. Children with nonpunitive teachers talk more about learning, achievement losses, and violations of school-unique values and rules.

Do children from nonpunitive teachers show more rational qualities in their responses? The answer to this question is not clear. Fairly direct attempts to measure this—codes for milieu likelihood of misconducts, for likelihood of consequences, and for appropriateness of consequences to the misconduct—did *not* show significant differences between the two groups. On the other hand, children with punitive teachers gave fewer abstract misconducts which result, in our camp study, was negatively correlated with age. But these same children also gave fewer reflexive justifications which result was positively correlated with chronological age. One interpretation of the findings that children of punitive teachers gave fewer abstract misconducts and fewer reflexive justifications is to regard these as indications of the unsettled and conflicted state of the attitudes regarding misconduct held by children with punitive teachers. When a child is inclined to misbehave but fears to, then a concrete act occurs to him—"hit George in the mouth"; when he is not pressed by his needs to misbehave, then an abstraction occurs to him—"be mean to people." Similarly, when he expresses this verbal act, a real consequence occurs—the target gets hurt or the perpetrator suffers a consequence; when he is not preoccupied with wrongdoing then a reflexive justification occurs to him—"it's not nice." A reflexive justification at this age may not be a primitive reply but a reflection of a settled issue: "You just don't do this because it's not nice."

Another interpretation is to regard the greater use of reflexive justification by the children with nonpunitive teachers as evidence of their greater trust and faith in school, i.e., of their internalization of school values more than children with punitive teachers. Inspection of the data showed the reflexive justification was used predominately in connection with rule violations (talking, running in halls, not taking seat, and the like). These misconducts are milieu-inconvenient which are disturbing to the milieu but which are without direct harm to either the actor or to others and do not violate an important moral code. Such misconducts to the first grade child have no real explanation except that "they're bad because they say so." As such, they express a sort of naive faith and trust in the rightness of what the teacher says.

Table 1: *A comparison of attitudes toward school misconducts held by children with punitive and nonpunitive first-grade teachers* ($N = 176$)

Misconducts and Explanations	% of Pu[a]	% of NPu[a]
I. Content and quality of the misconducts:		
A. Physical assaults on others	38	17
B. Milieu-serious misconducts	89	63
C. Coder-serious misconducts	48	27
D. Abstract misconducts	27	52
II. Content and quality of the explanations:		
A. Peers as objects of consequences	94	61
B. Physical damage to objects of consequences	60	23
C. Serious harm to others	45	18
D. Reality-centered retributions	21	48
E. "Reflexive justifications" as explanations	11	26
III. Role of self in misconducts:		
A. Ego-alien misconducts	26	11
B. Premeditated misconducts	29	15
IV. Aggression:		
A. Overall aggression ("blood and guts")	49	24
V. Concern with school-unique objectives:		
A. Learning and achievement losses	20	43
B. Institutional law violations	49	62

Note.—All differences in percentages are significant at the .05 level or beyond.

[a] *Pu stands for those children who have punitive teachers; NPu refers to those children who have nonpunitive teachers.*

SUMMARY

Three pairs of punitive vs. nonpunitive first grade teachers were selected from three elementary schools. The 174 children in these teachers' classrooms were individually interviewed about what they thought was "the worst thing to do in school" together with their explanations of why these misconducts were bad. Regarding their responses as expressions of their preoccupations it was concluded that, as compared with

children who have nonpunitive teachers, children who have punitive teachers: manifest more aggression in their misconducts, are more unsettled and conflicted about misconduct in school, are less concerned with learning and school-unique values, show some, but not consistent, indication of a reduction in rationality pertaining to school misconduct. A theory that children with punitive teachers develop less trust of school than do children with nonpunitive teachers was also presented to explain some of the findings.

REFERENCES

GUMP, P. V., & KOUNIN, J. S. Sex and milieu influences in children's concepts of misconduct. *Amer. Psychologist,* 1959, *14,* 346. (Abstract)

KOUNIN, J. S., & GUMP, P. V. The ripple effect in discipline. *Elem. sch. J.,* 1958, *59,* 158–162.

32

The Relationship Between Pupil Self-Evaluation, Teacher Evaluation of the Pupil, and Scholastic Achievement

Ira J. Gordon and Patricia C. Wood

Gordon and Wood explore another aspect of the pupil-teacher relationship—the interaction between the teacher's and the pupil's perceptions of the pupil's ability. This study is illustrative of the ability of a classroom teacher to carry on effective research with assistance from a university resource person.

Self theorists have proposed a fundamental relationship between self-concept and behavior. In addition, they have postulated that threat to the self decreases the ability of the person to perceive both self and situation accurately.

Previous research leaves unresolved the relationship between teachers' behavior and the self-concept of individual pupils. Although it is a widely held belief among educators and many psychologists that such a relationship exists, empirical data are needed. Perkins (1958) and Staines (1958) both found relationships between teacher behavior and self-conceptualization in children. Sears (1960) found that there was a relationship between pupil self-ratings on mental ability and teacher ratings of mental ability on youngsters in middle childhood.

The research on accuracy of self-evaluation by children suggests first, that there is a wide range in ability to view self realistically (Brandt, 1958), second, that there seems to be a tendency to overestimate (Wylie, 1961) and third, that accuracy seems to be a function not only of the variable being estimated (Wylie, 1961) but also the degree of "threat" (Perkins, 1958). A fourth generalization from the research has been stated by Wylie (1961): that there is some limited evidence of individual consistency of self-rating, but a lack of detailed information.

This study is designed to explore the nature of the agreements, or disagreements, between pupil and teacher estimates of the pupil and the relationships of these evaluations to scholastic achievement. What happens when teacher and pupil agree? Is the pupil then more able to evaluate his performance realistically? Is teacher-pupil disagreement related to more perceptual distortion on the part of the pupil? Theoretically, there should be a relationship between the teacher's and the pupil's estimate. Is this borne out empirically?

Further, are the pupils in this study individually consistent? Does the youngster who overestimates himself on a general self-rating scale on a variety of school items also overestimate himself on his estimate of performance on a specific task? Is there a general tendency, throughout the group, to overestimate?

PROCEDURE

Three classes of fifth graders and two classes of sixth graders (72 boys and 76 girls) in a metropolitan Florida school and their five teachers served as subjects. These teachers averaged 15 years of experience, were all women, two held master's degrees, while the other three held bachelor's degrees in elementary education.

In the spring of 1960, the pupils rated themselves on the 15 school items of the "How I See Myself" 5-point scale developed by Gordon from the Jersild (1952) categories. The following are a sample of the fifteen categories:

From *Journal of Educational Research,* 1963, *56:* 440-443

In school, I don't stay with things and finish them.	1 2 3 4 5	In school, I stay with something till I finish
I don't do well in school work	1 2 3 4 5	I do well in school work
School isn't interesting to me	1 2 3 4 5	School is very interesting to me
I'm not as smart as the others in my class	1 2 3 4 5	I'm smarter than most of the others in my class
I don't read well	1 2 3 4 5	I read very well

The children were given the following explanation of the questions:

"I would like for you to help me with a study I'm making of teacher-pupil relationships—the way you and your teacher work and study together. These sheets are for my use only. None of your classmates, or your teacher, or the principal, may see them—this is just between you and me.

"There are fifteen items on this sheet that I want you to mark according to the way you feel most of the time. No one feels the same way about school all the time, but try to think of the way you feel most of the time. Please remember that no one else will see your papers."

The children were given time to ask questions and then each item was read, and the child circled the number that represented his feeling on the item.

In order to avoid "set," about half the items randomly distributed were presented with one as the positive answer, and five the negative. For scoring purposes, all positive items were ranked as five. A total score on the fifteen items was computed for each child. The children were listed from highest to lowest using this score, then this list was divided into stanines.

Each teacher was asked to rank his pupils from highest to lowest on the same self-report scale, then this list was divided into stanines. The two lists were compared. If the teacher placed the child in the same stanine as he placed himself, they were said to agree. If the teacher did not place the child in the same stanine as he placed himself, they were said to disagree.

Two weeks later, the children were asked to rate themselves as to their estimate of their relative position in class achievement, as measured by the Stanford Achievement Test. They were told that in a few weeks they would be given the regular yearly achievement tests. They were asked to consider themselves in relation to their classmates—"the number of people in the room you think are smarter than you, and then the number of people you think are not as smart as you are generally." They were asked to write on the paper given them the number that represented their class position, whether they thought they were first, or second, or fourth, or sixteenth, or thirty-fourth. These estimates were arranged in order from highest to lowest in each class; then each list was divided into stanines.

The teachers were asked to rank their classes from highest to lowest on achievement as measured by the Stanford Achievement Test. The lists were divided into stanines and compared to the child's estimates. The same criterion as before was used for agree and disagree.

Two weeks later, each teacher gave the intermediate battery, Form JM, of the Stanford Achievement Test. The scores of the pupils were listed from highest to lowest and divided into stanines. If the teacher or the child placed the child in the same stanine on self-report scale or achievement estimate as his actual achievement test score, he (the teacher or the child) was said to be accurate. If he did not place the child in the same stanine as the achievement test score, he was said to be inaccurate. In cases where the teacher and the child disagreed, either one could be accurate by placing the child in the same stanine as his achievement test score.

For example, if the child fell in the third stanine on the self-report scale and the third stanine on his actual achievement test score, he was said to be accurate on his self-report. The same definition applied to test estimate.

The hypotheses to be tested were:

1. There will be a positive relationship between the child's agreement with the teacher on the self-report scale and his accuracy on the self-report scale as an estimate of achievement.
2. There will be a positive relationship between the child's agreement with the teacher on estimate of test ranking and his accuracy in ranking himself.
3. When the teacher and child are inaccurate, they will agree on the direction of inaccuracy (over- or under-estimate) in comparing self-report scale to actual achievement test rank.
4. When the teacher and the child are inaccurate on the achievement estimate, they will agree on the direction of the inaccuracy.
5. When the teacher and the child disagree on the self-report scale, the teacher's ranking will be closer to actual ranking on achievement test.

6. Where the teacher and child disagree on the estimate of achievement, the teacher's ranking will be closer to actual rank on achievement test.
7. Children who err in their estimate on one measure will err in the same direction on the other measure.

RESULTS

The data for all the hypotheses, except the fifth and sixth, were analyzed by means of the chi-square test of significance.

Table 1 presents the findings on the teacher-pupil hypotheses in which chi-square was used. The first hypothesis, that there will be a positive correlation between agreement on self-report and the child's accuracy on ranking himself in the same relative class position on the self-report scale that he holds on achievement test performance, was upheld. The number who were accurate was small, but of those who were, twice as many also had teacher-pupil agreement.

High accuracy was not expected for two reasons. First, agreement on a stanine classification is a rigorous test. In a class of 27 pupils, the stanine distribution is: 1, 2, 4, 5, 7, 5, 4, 2, 1. The child could be ranked second on self-report and actually be third in achievement and be classified as "inaccurate." Second, the child was rating himself generally on many school items, including attitudes toward school subjects and teachers as well as ability. This rating was being used against test achievement on a standardized test.

When the children were asked to rank themselves on the particular task, the number who were inaccurate doubled but was still less than one-third of the group. There was no relationship between agreement and accuracy on the specific task.

Since teacher and pupil both tend to be inaccurate by this rigorous definition, is there any pattern to their inaccuracy? Three types of hypotheses were explored, one dealing with direction of error, the second dealing with amount, and the third with consistency of error. Hypotheses three and four (direction of error) were upheld (see Table 1). We cannot assign causation, but the high degree of relationship is noteworthy.

Hypotheses five and six (amount of error) were tested by the use of the *t* test for correlated means. The mean difference between the discrepancy scores (actual stanine–estimated stanine) for both pupils and teachers on both self-report and test estimate were computed. In both cases, the teachers were closer to the actual rank than were the pupils. The differences between teacher and pupils were significant at the .001 level. That is, even though the teachers missed the correct rank, they were closer to being "on target" than were the pupils.

Table 1: *Teacher and child estimates and scholastic achievement rating*

Variable	Number of Children		
	Accurate	Inaccurate	Chi-Square
1. Agreement between teacher and pupil on Self-report			
Agree	10	26	
Disagree	5	105	13.46*
2. Agreement between teacher and pupil on achievement estimates			
Agree	13	29	
Disagree	21	83	1.38
	Overestimate	Underestimate	
3. Direction of error of estimate, Self-report			
Teacher overestimates	25	14	
Teacher underestimates	10	32	19.06*
4. Direction of error of estimate, achievement			
Teacher overestimates	25	12	
Teacher underestimates	10	25	16.25*

** Difference is significant at the .01 level of confidence.*

Table 2: *Consistency of children's estimate between Self-report and test estimate*

Direction of error of estimate on rank achievement test	Direction of error, Self-report, to actual rank on achievement		Chi-Square
	Overestimate	Underestimate	
Overestimate	40	11	34.6**
Underestimate	11	42	

** *Difference is significant at the .001 level of confidence.*

The seventh hypothesis, that pupils would be consistent in their over- or underestimation was upheld (see Table 2). The child who overestimated his achievement test score also tended to overestimate himself on the self-report scale.

The data were analyzed for sex differences, but none were found. Teachers were neither more nor less accurate on boys than girls; neither more nor less in agreement with one sex or the other. Neither boys nor girls had higher accuracy, nor different tendencies to over- or underestimate.

DISCUSSION

In this study, teachers were not able to predict with much success the way children would evaluate themselves on self-reports about school and estimate of test performance. There was agreement on direction of error of evaluation, but little on actual ranking. It may be that the teacher's evaluation of the child influences the way he evaluates himself. We seem to have an "over-reaction" situation. The teacher overestimates the child a little, the child overestimates himself even more. Although we cannot state a cause-effect relationship, Staines' (1958) work would suggest this possibility.

The relationship between agreement and accuracy on the self-report tends to support the idea that agreement might be threat-reducing leading the child to see himself realistically. Disagreement may lead to the tendency for the pupil to distort, hence the over-reaction.

Only seven children were accurate on both estimates. Three of these had complete agreement with the teacher, and three more agreed with the teacher on one estimate. Of the eight others who were accurate on self-report but either over or under-evaluated their test estimate, none had complete agreement with the teacher, three had no agreement, and only two agreed with the teacher's estimate of achievement. Although these numbers are too small for any statistical meaning, they do reflect a possible trend of a positive relationship between pupil-teacher agreement and pupil accuracy of self-estimate.

We found this group of pupils to be consistent in both under- and over-estimation. We found no more tendency to overestimate than to underestimate. Further, if we couple this pupil consistency with the significant relationship to teacher estimate, there is a pattern of consistent error by both teacher and child. Although the magnitude of the teacher's error is significantly less than the pupils', the consistency of direction seems most significant. The big question, of course, is who is influencing whom? Is it that the teacher understands how the child sees himself, or does the teacher's evaluation shape the child? We cannot say from this study. We can say that the child's general view of self in relation to school is reflected in his specific view of his ability on a standardized achievement test.

SUMMARY

Teachers and children in five fifth- and sixth-grade classes were asked to estimate the child's performance on a standardized achievement test. The children also rated themselves on a "How I See Myself" scale on general feelings about themselves in relation to the school situation. The teachers estimated the rank order of the children on these reports. All ranks were placed in stanines, and the amount of agreement between teacher and child, and degree and amount of agreement between self-reports, estimates, and actual test rank were analyzed. Relationships were found between pupil-teacher agreement on self-reports and accuracy of the self-report as an estimate of actual test rank and between direction of error on both self-report and test estimate between teacher and pupils. The pupils were consistent in the direction (over- or underestimation) on both measures. Teachers were closer to correct estimates of test ranks than were pupils. Teacher-pupil disagreement seems to be related to perceptual distortion in the same direction by both. The role of threat was discussed as a possible factor in the amount and degree of pupil error.

REFERENCES

BRANDT, R. M. The accuracy of self estimate: A measure of self-concept reality. *Genet. Psychol. Monogr.*, 1958, *58*, 55-99.

JERSILD, A. T. *In search of self.* New York: Bureau of Publications, Teachers College, Columbia University, 1952.

PERKINS, H. V. Factors influencing change in children's self concepts. *Child Develpm.*, 1958, *29*, 221-230.

SEARS, PAULINE. The pursuit of self-esteem: The middle childhood years. *Newsletter*, Division of Developmental Psychology, 1960, 1-5.

STAINES, J. The self-picture as a factor in the classroom. *Brit. J. educ. Psychol.*, 1958, *28*, 97-111.

WYLIE, R. *The self-concept.* Lincoln: University of Nebraska Press, 1961.

33

The Effect of a Teaching Program in Behavioral Science on Changes in Causal Behavior Scores

Ralph H. Ojemann and Bill C. Snider

For nearly two decades Ojemann has been guiding the research at the University of Iowa on the role of guided learning experiences in changing children's concepts of behavior. This paper is one of the series of reports on the work. The general construct is highly sophisticated and difficult to test empirically. Studies such as this are essential in order to increase our knowledge of the transactional relationship between the growing child and his cultural environment. See also Ojemann and Pritchett (24) and Crabtree (35).

The purpose of this paper is to describe the changes in the observed daily behavior of the child that appeared following the administration of a teaching program in "behavioral science." This investigation is part of the Preventive Psychiatry Research Program of the University of Iowa. One of the aims of this program is to develop a sequence of learning experiences beginning in the earliest years to help the child acquire an understanding and appreciation of the dynamic nature of human behavior; that is, to develop a causal approach to the social environment and to test the effects of such experiences.

The nature of the learning program has been described elsewhere (Ojemann *et al.*, 1955) and various investigations of effects have been reported in previous papers. This paper is concerned with changes obtained on a behavior observation form used by trained observers. The observers were not part of the teaching team and the data were obtained without the child's knowledge of the purpose or nature of the observations.

The investigation involved the development of a method for making and scoring observations of the child's daily behavior and its use with experimental and control groups. The development of the instrument used in making the observations will be considered first. This will be followed by a description of the subjects, the experimental design, and the analysis of results.

DEVELOPMENT OF THE OBSERVATION FORM

In constructing the instrument for making the observations, use was made of the results reported by Kooker (1951) in his study of an observational method for measuring insecurity and boredom in school age children. One of the difficulties in the use of observational scores obtained from observations of the daily interactions of children with their environment is that the meaning of a given behavior incident depends in part upon the situation in which it occurs. For example, crying after a severe fall usually does not have the same meaning as crying where there are no physical injuries. Kooker found that when behavior items were carefully defined as to the situation to which they were to be considered pertinent and when several degrees of frequency were available for rating the behavior,

---

Scheduled for publication in *Journal of Experimental Education*

the reliability of observational ratings prepared by observers with but a small amount of training ranged from .63 to .86.

An effort was made to develop behavior items which could be clearly defined as to the setting to which they applied and which permitted behavior varying from noncausal to causal in character. For example, the first item on the final observation form reads, "If a situation involving some difficulty between two or more people arises and has to be dealt with, the child suggests or otherwise shows that he expects the class or the teacher to investigate the causes of the problem or difficulty." This is defined as referring to situations that arise during the day in the classroom or on the playground and come up for discussion between teacher and one or more pupils or between two or more pupils. Thus, this item refers to situations that have to be dealt with and where there is opportunity for discussion by the pupils being observed.

Furthermore, the pupil is free to behave as he wishes. It is not a class exercise on which he is graded. Also, there are no special skills required for the performance of the behavior itself.

Items were selected which involved behavior indicative of a sensitivity or lack of sensitivity to causes, alternatives, and probable consequences. Each item was rated for each subject on a four point scale; namely, frequently, fairly often, seldom, and never.

Additional examples of items may be helpful at this point to give a further indication of their nature. Some items refer to situations in which a child meets a difficulty for which alternatives clearly within the limits of the child's ability are available. For example, item #3 reads, "The child keeps on trying (persists) when he meets a difficulty in his learning activity." This refers to any learning situation in which the child meets some kind of obstacle. He can attempt various hypotheses in solving the problem. He can try. If he finds it beyond his ability, he can seek help from other sources such as books, his classmates, and the teacher. On the other hand, he can give up, show no interest in finding out more about the cause or nature of the difficulty or available alternatives.

The reasoning underlying the selection of such an item was as follows: If the child were applying the idea that when one meets a difficulty there are causes which can be inquired into and which upon inquiry may suggest alternatives, he would be more likely to keep on trying as defined above than "giving up." It is to be noted that this item specifies that the situation be one for which alternatives within the limits of the child's ability are available.

Similar reasoning governed the selection and inclusion of items involving the destructiveness of materials for no obvious reasons, overdependence on the teacher, withdrawal, or overaggressiveness when meeting a difficulty.

Another type of item refers to the "labeling" or "prejudging" of others. This refers to negative or unfriendly remarks such as "Speedy" when the child is slow; "Brain" when the opposite is true or "Teacher's Pet." It does not refer to common nicknames which the child has acquired. It was reasoned that the child who tries to understand is much less likely to label. It is more difficult to label when one begins to consider various hypotheses as to how a child may have become "clumsy," "show-off" or "slow." The individual when understood usually turns out to be more complex than a label suggests.

A total of 17 items comprised the final observation form. An original group of twenty-five items was submitted to three judges thoroughly familiar with the meaning of causal behavior. They were asked to select the items that clearly involved causal-noncausal behavior. The twenty items selected by all three judges were tried out with a group of three observers for clarity of meaning and suitability for observation in the school situation. Considerable care was exercised in clarifying the meaning of each item. Seventeen items were judged by the observers as meeting the criteria of clarity and suitability for observation. These seventeen items were used in the final observation form.

The four levels of frequency used in rating the child on each item were given numerical values of 4, 3, 2, 1 with the higher value representing the more causal end of the continuum. Greater refinement of scaling was not attempted at this time since it was considered desirable to obtain more experience with the items in both experimental and control subjects.

PROCEDURE IN MAKING OBSERVATIONS

The observers were mature females who had a college degree but who were not members of the school staff and hence did no teaching. They were asked to spend one week in the classroom in the fall and one week again in the spring. The period of one week was selected after preliminary tryout to determine the length of time an observer would need to get acquainted with the children in the class and to make the ratings. The one week period was preceded by visits to the class for a brief period to meet the teacher.

The observer was introduced to the class as a visitor. The occasional presence of a visitor was not unusual since the school system cooperated in a teacher training program.

Neither the class nor the teacher were aware of the detailed items on the observation form. The teachers in the experimental classes were told before they entered the experimental program that various tests and other types of measurements would be made from time to time and that the detailed nature of the data gathered would not be discussed with them. Since the experimental teachers had worked with the investigators in past years, they accepted the observers without further question.

The control teachers also understood that various tests and other types of measurements would be made and to meet the requirements of careful research the detailed nature of the measurements could not be discussed. Each teacher was assured, however, that all data obtained by the research workers would be used for research purposes only, would not be available to the administrators of the school system, and would not be used in any way for promotion or other related purposes. This was in accord with a continuing agreement with the school system in which the investigation was carried out and efforts were made to provide the teacher with continuing assurance.

The observers were given preliminary training in the use of the observation form. This was done in the spring preceding the year in which the observations were made. This preliminary training consisted of a group discussion of the items on the observation form, preliminary observation of three children by each observer, and subsequent group discussion and clarification of the items. Considerable care was taken in the discussion to consider all questions raised by the observers.

After this preliminary training, a study of reliability was made. Simultaneous observations were made by observers of one class at the fourth and one at the fifth grade level. The correlation of the behavior scores obtained by observers I and II, using a class of 32 fourth grade subjects, was .69. The corresponding correlation of observers II and III, using a class of 28 fifth grade pupils, was .67. In the final study, observers I and II gathered the data at the fourth grade level; II and III at the fifth grade level. Each observer obtained data from two experimental-control pairs.

The observers indicated the ratings on the individual items. They did not score the items and the scoring plan was not discussed with them. They did not know what numerical values would be assigned to the various choices which each item presented. This procedure was followed to increase the difficulty for an observer to recall in the spring of the year the ratings given individual children on the different items the previous fall. In the judgment of the investigators, this procedure was rather effective.

SUBJECTS

The subjects of this study were the children in four fourth grade experimental classes and their corresponding controls and four fifth grade experimental classes and their corresponding controls. The investigation was carried out in a midwest industrial city of approximately 90,000 population.

Each experimental class was in charge of a teacher who had received training in teaching the dynamics of behavior and had available appropriate content which he had helped to prepare.

Each experimental class was paired with a control in the same school system. In selecting the controls, an attempt was made to choose a class having a teacher of the same sex, age, and years of training as the corresponding experimental teacher and in which the pupils were judged to have a socio-economic background comparable to that of the experimental class. This procedure of obtaining control groups was used in previous studies. The initial observations were made in the fall, beginning two weeks after school had started. The final observations were made in April and May.

Two measures, I.Q. scores and initial behavior scores, were studied to determine the equivalence of the experimental and control groups. Table 1 gives the data relative to I.Q. and initial behavior scores for the four fourth-grade E-C pairs. Inspection of the data relative to the I.Q. scores indicates that there were no significant differences between the individual E-C pairs or between the total E-C group with respect to this measure.

An examination of the data relative to the Initial Behavior Scores shows one significant difference between the experimental and control groups. The experimental group of the pair E-C-4-4 scored significantly higher than the control group. The total fourth grade group showed no significant difference in initial behavior scores.

Corresponding data for the fifth grade are presented in Table 2. Inspection of this table reveals one significant difference between E-C pairs in I.Q. scores; namely, E-C-5-4. The difference is significant at the 1% level of confidence. The difference between the total fifth grade experimental and control group was not significant.

An analysis of the initial behavior scores reveals two significant differences between E-C pairs; namely, E-C-5-2 and E-C-5-4. In both cases, the control group scored significantly higher than the experimental group. The total fifth grade group showed no significant differences in initial behavior scores between experimental and control groups.

In the analysis of results, the changes from initial to final scores will be used. This will take account of differences in initial behavior scores. The "final minus initial" behavior scores for E-C-5-4 will be adjusted by co-variance to take account of the significant difference in I.Q. between the experimental and control groups.

RESULTS AND DISCUSSION

The results will be reported for each grade separately and then brought together. The data for the four fourth grade experimental classes and their respective controls are given in Table 3. Included in the tabulations are the number in each class, the mean of the "final minus initial" behavior scores for both experimental and control groups and the corresponding *t*. Similar data are given in the last line of the table for the total group.

An examination of the data in this table indicates that two of the individual experimental classes revealed significant gains over their respective controls and two did not. The difference for the total fourth grade group (106 experimental and 100 controls) is significant at the 5% level of confidence. Thus, it appears that when the total fourth grade group is considered, the learning program produced a significant difference in behavior as measured by observational scores but the results vary somewhat among the individual classes. There was a tendency toward a difference in favor of the experimental group in all classes. These data suggest that there may be differences in the effectiveness of the individual teachers in producing behavioral changes.

A further examination of the "final minus initial" behavior scores for the control group reveals an interesting phenomenon. Three of the four control groups yielded scores which are more "arbitrary-judgmental-noncausal" as defined in this study at the close of the experiment than at the beginning. A similar result appeared in only one of the experimental classes. This tendency for children under ordinary conditions as they prevail at present in school to move in the noncausal direction was also observed in the "paper-and-pencil" Problem Situations Test used in a previous study (Ojemann *et al.*, 1955). Since it appeared in a previous study in which a paper-and-pencil test was administered to children and in the present investigation in which scores were obtained using independent observers, it is very probable that this finding is not an artifact arising from the method used in the present study. It appears more fundamental. It is the type of result we might expect from the analysis of the noncausal nature of the content in general readers, history and related areas presently taught in schools which was described in a previous paper (Ojemann, 1958).

We might also expect this finding when we consider arbitrary judgmental tendencies often observed in the behavior of the untrained teacher. . . .

The data for the four fifth grade E-C pairs and the total fifth grade are presented in Table 4. It will be recalled from the data in Table 2 that in one of the E-C pairs (E-C-5-4) the experimental group differed significantly in I.Q. from the control group. In this class there was a correlation of .41 between gains and I.Q. scores. Accordingly adjusted mean changes are reported in Table 4 for the E-C-5-4 subjects. Three of the experimental classes made significant gains over their respective controls. The difference for the total fifth grade group (99 experimental subjects and 113 controls) is significant at the 1% level of confidence.

The phenomenon of loss in causal scores is again observed in the case of the control groups for the fifth grade.

There was also a tendency for the results at the fifth grade level to be somewhat more consistent. Three of the four E-C pairs showed a difference significant at the 1% level of confidence. In the fourth grade, one of the four individual E-C pairs showed a difference significant at the 1% level of confidence and one at the 5% level.

At both the fourth and fifth grade levels the total experimental group yielded significantly

Table 1: *I.Q. scores and initial behavior scores—fourth grade*

Group	N		I.Q.					Initial Behavior Scores				
	Exp.	Cont.	Mean-Exp.	Mean-Cont.	Diff.	*t*	Sig.	Mean-Exp.	Mean-Cont.	Diff.	*t*	Sig.
E-C-4-1	22	21	107.05	110.43	3.38	— .87	*ns*	40.64	45.62	4.98	—1.52	ns
E-C-4-2	24	25	109.12	104.00	5.12	1.39	*ns*	41.62	41.52	.10	.03	ns
E-C-4-3	31	25	112.55	110.48	2.07	.56	*ns*	50.26	48.92	1.34	.77	ns
E-C-4-4	29	29	120.07	124.79	4.72	—1.21	*ns*	50.55	45.72	4.83	2.79	1%
Total 4th	106	100	112.69	113.00	.31	— .15	*ns*	46.39	45.45	.94	.72	ns

Table 2: *I.Q. scores and initial behavior scores—fifth grade*

Group	N		I.Q.					Initial Behavior Scores				
	Exp.	Cont.	Mean-Exp.	Mean-Cont.	Diff.	*t*	Sig.	Mean-Exp.	Mean-Cont.	Diff.	*t*	Sig.
E-C-5-1	21	26	103.62	106.50	2.82	—1.18	ns	43.14	42.73	.41	.13	ns
E-C-5-2	24	26	106.12	104.11	2.01	.53	ns	47.75	52.36	—4.61	—2.42	5%
E-C-5-3	22	29	103.95	109.62	5.67	—1.88	ns	53.77	53.83	— .06	— .03	ns
E-C-5-4	32	30	116.06	106.53	9.53	4.61	1%	52.09	56.10	—4.01	—3.02	1%
Total 5th	99	113	108.32	106.72	1.60	1.06	ns	49.52	51.51	1.99	—1.70	ns

Table 3: *Differences between final and initial behavior scores—fourth grade*

4th Grade	N		Final Minus Initial Scores			
	E	C	Mean E	Mean C	*t*	Sig.
E-C-4-1	22	21	3.23	—3.14	3.44	1%
E-C-4-2	24	25	1.92	— .64	2.13	5%
E-C-4-3	31	25	—2.35	—2.96	.35	ns
E-C-4-4	29	29	2.75	1.93	.53	ns
Total 4th	106	100	1.16	—1.00	2.33	5%

Table 4: *Differences between final and initial behavior scores—fifth grade*

5th Grade	N		Final Minus Initial Scores			
	E	C	Mean E	Mean C	*t*	Sig.
E-C-5-1	21	26	— .71	— .81	.04	ns
E-C-5-2	24	26	3.38	—1.00	3.86	1%
E-C-5-3	22	29	— .86	—5.14	3.15	1%
E-C-5-4	32	30	2.27*	—3.78*	4.58	1%
Total Fifth	99	113	1.32	—2.86	5.46	1%

* *Adjusted for difference in I.Q. scores. The test for homogeneity of variance was satisfactory.*

higher causal behavior scores as defined in this study than the corresponding control group.

It is quite probable that a further refinement in the instrument for recording the observations will produce higher interobserver reliability and thus sharpen the differences. It is also highly probable that improvements in the learning program can bring about greater changes in the experimental groups. Studies of growth in causal knowledge and related attitudes have indicated that lengthening the time during which the learning program operates (Muuss, 1960) produced increasing changes in the knowledge and attitude dimensions.

The data also suggest that there may be differences in effectiveness among teachers and one of the problems for further investigation is that of identifying some of the factors that enhance or reduce the effectiveness of the teacher in developing causal behavior in pupils.

SUMMARY

The purpose of this investigation was to study the changes in observed behavior that appeared following the administration of a teaching program in behavioral science. The subjects of the study were 206 fourth grade pupils (106 in experimental groups and 100 in control groups) and 212 fifth grade pupils (99 in experimental groups and 113 in control groups). The investigation was carried out in a midwest city of approximately 90,000 population. Observations of the child's daily behavior were made by independent observers using an observation report form developed for this investigation. The interobserver reliability varied from .67–.69.

The results of the investigation indicated that the total fourth grade experimental group made a significant gain at the 5% level of confidence over the control group. A further analysis of the data at the fourth grade level indicated that two of the four experimental classes comprising the total group showed significant gains over their respective controls and two did not. The differences in all four experimental groups over their respective controls were in favor of the experimental groups.

The findings at the fifth grade level indicated that a significant gain was also made by the total fifth grade group over its control group. This difference was significant at the 1% level of confidence. A breakdown of the data at this level into comparisons of the four individual experimental classes with their respective controls indicated that three of the four experimental-control comparisons yielded a significant difference.

The learning program produced significant changes at both the fourth and fifth grade levels. The variations among the individual classes suggested a study of the factors producing differences in teacher effectiveness.

REFERENCES

KOOKER, E. W. An investigation of security, insecurity, achievement, and boredom in elementary school children. Unpublished doctoral dissertation, State University of Iowa, 1951.

MUUSS, R. E. The effects of a one- and two-year causal-learning program. *J. Pers.*, 1961, *28*, 479–491.

OJEMANN, R. H., LEVITT, E. E., LYLE, W. H., JR., & WHITESIDE, M. F. The effects of a "causal" teacher training program and certain curricular changes on grade school children. *J. exp. Educ.*, 1955, *24*, 95–114.

OJEMANN, R. H. Basic approaches to mental health: The human relations program at the State University of Iowa. *Personnel guid. J.*, 1958, *37*, 198–206.

34

Teaching Gifted Elementary Pupils Research Concepts and Skills

E. Paul Torrance and R. E. Myers

Torrance has stimulated many studies of creativity, including those reported in his book Guiding Creative Talent *(1962). As a part of this effort, he has focused on gifted children and on the task of examining what types of concepts they can learn far sooner than they would normally experience them. His work parallels*

From *Gifted Child Quarterly*, 1963, 1–6.

Ojemann's (24 and 33) in that it introduces guided learning experiences along with careful evaluation.

The teaching of research concepts and skills has usually been reserved for the graduate-school period. Discussions on teaching elementary pupils how to do research are usually limited to procedures for "looking something up in the library" and almost never touch upon the concepts and methods of creative scientific research. Textbooks on methods of teaching in high school and in the undergraduate college are seldom much better. Yet, important discoveries have been made by elementary pupils, high school students, and undergraduates. Louis Braille at the age of 10 conceived of his idea for an alphabet for the blind and by 15 had worked out the details. Many of the world's most outstanding medical discoveries were made by undergraduate students or were begun during the discoverer's undergraduate days.

The senior author has for some time believed that most research concepts and many research skills could be taught to elementary pupils, especially gifted ones. It was his belief that if gifted children could be taught these concepts and skills at an early age, they would have available some powerful tools to aid them in their thinking. This should make learning more exciting and the search for "truth" more rewarding. For the past three years, he and his staff have had the opportunity to test this idea in part through a five-day course taught to the High Achievers Class at the Riverside Elementary School of Bloomington, Minnesota.

Each year, the course has been varied. In many ways, the 1961–62 course was a better balanced one than previous courses and will be described briefly. It may suggest quite different ideas among some workers. Some may want to expand the program as the authors developed it or select from it.

The 1961–62 class consisted of 46 pupils selected on the basis of achievement and intelligence (minimum Stanford-Binet IQ, 135). Each morning these pupils function as two separate groups under different teachers (Gilbertson, 1960). During this period, they follow essentially the same curriculum as other sixth-grade classes, with the exception that they have lessons in French. This leaves the afternoons free for other ventures. On alternate weeks they work on their "strengths and weaknesses" under the direction of their teachers. During other weeks, both groups study for a week with "Alter Teachers" who are expert in some special field such as: chemistry, journalism, dentistry, music, psychology, architecture, law, and the like. The course on Educational Research was conducted as a part of the alter-teacher program. . . .

OBJECTIVES

The major objectives of the course on Educational Research were:

1. To familiarize gifted elementary pupils with some of the most powerful concepts of research in the behavioral sciences, in order that these concepts might become tools in the learning and thinking early in their school careers.
2. To communicate to gifted pupils the excitement of doing original research, exploring the unknown and pushing forward knowledge.
3. To provide experiences in participating in educational research and in conducting experiments with real consequences.
4. To develop in gifted elementary pupils some skills in formulating hypotheses, testing them, and in reporting the results.

WHAT IS THE RESEARCH PROCESS?

On the first day, the children learned that there are many things which teachers and other educators do not know about how to help boys and girls to learn most effectively. They learned that one of the ways of finding out about these unknowns is through research. The youngsters offered their own ideas about research—ways of digging into things, looking things up, experimenting, etc. They learned that research involves a more or less formal, systematic, intensive process of carrying on the scientific method.

In quick order, the sixth-grade pupils experienced some of the research process through the "What's in the Box?" game. The two classes which comprise the High Achievers at Riverside School were separated, one group working under each of the authors. In each case the group was told that the object in the box was round. The group then made some wild guesses of round things which might be in the box. They then wanted to experiment—lifting, shaking, smelling, and otherwise examining. One group found that the object made no sound and immediately eliminated all objects which would make a sound, which were rather heavy, and the like. The other group found that the object in its box made a sound, was light, and was apparently quite small. This group then eliminated soft objects, heavy objects, and large objects. Both groups at this point made some more guesses and started asking questions which could be answered by "yes" or "no." One by one, they eliminated the guesses they had made and finally guessed the correct identity of the object—in one case, a powder puff; in the other, a Sucret (cough drop). This was then verified.

The person who identified the object then described the process of *synthesis* by which he arrived at the correct answer. From this, the class identified the following steps in the process:

1. Recognizing that you don't know something.
2. Making wild guesses on the basis of available facts.
3. Experimenting, testing, eliminating guesses.
4. Making better guesses.
5. Asking questions to test hypotheses (guesses).
6. Eliminating bad guesses on the basis of additional information.
7. Making better guesses, synthesizing facts obtained.
8. Verifying the final answer.

THREE KINDS OF RESEARCH

The High Achievers then learned that there are three major kinds of research and that the concepts learned in the "What's in the Box?" game are applied to each of them. They grasped easily the essential characteristics of historical, descriptive, and experimental research and learned that they would explore each firsthand during the week.

Historical Research

Historiography and the process of doing historical research were introduced first. Right in class, the pupils were given the first of two "How Did You Grow?" exercises which required them to plot the trends in their own development from first through sixth grade in shoe size, height, curiosity, reading speed, imagination, spelling ability, independence in thinking, and arithmetic computation.

The first day ended with a real challenge to the youngsters—they were invited to engage in some actual historical research. The High Achievers were given a set of the same charts they had just filled out and were told to collect all of the data they could to test the trends which they had already hypothesized.

On the next day the class reported dozens of sources they had used: parents, teachers, brothers and sisters, other relatives, photograph albums, report cards, old clothing and shoes, baby diaries, and the like. They agreed that these could be classified either as records or witnesses. The children found that they had been introduced firsthand to the problem of *bias* and of *evaluating evidence* and reaching a *conclusion*. They found that even their parents perceived their development in vastly different ways and guessed why.

Descriptive Research

In order to give some experience with a descriptive study, the High Achievers were administered the Ask-and-Guess Test, using the picture of Old Mother Hubbard. First, they asked as many questions as they could about the picture and the events occurring in it. Then they spent five minutes making guesses about causes and another five minutes guessing consequences. After they had counted the number of questions and recorded them on the blackboard, they decided that one way of using these data to describe the group would be to find the average or mean. They noted that the girls had a larger mean (12.17) than the boys (9.43) but could not tell whether or not such a difference was just an accident—a matter of chance. They were introduced to the concept of levels of confidence and the t test. It was found that this difference could be accepted at about the two per cent level of confidence; that is, it would fail to occur in only about two cases in one hundred.

In a similar manner, the concepts of the median, the mode, variance, and the like were developed.

Experimental Research

On the third day, the High Achievers served as subjects of an experiment on the effects of differential rewards upon creative thinking. To teach the concepts of randomization, experimental treatments, and the like, the two classes were divided randomly into Groups A and B. This was done by means of drawing assignment cards which had been carefully shuffled. Torrance and Myers tossed a coin to see who would have Group A and who would have which treatment. A prize of two dollars was offered to the pupil in Group A who could think of the largest number of ideas, regardless of quality; and a prize of twenty-five cents was offered to the pupil who could think of the largest number of original ideas. The treatment for Group B was just the reverse—the pupil with the most original ideas was offered a prize of two dollars, while the one with the largest number of ideas, regardless of quality, was offered twenty-five cents. The task was to think of ideas for making a toy dog more fun to play with. Ten minutes were allowed for the task. A response was scored as original if it had been given by fewer than five per cent of the subjects in the norm population for the Minnesota Tests of Creativity. Responses were also scored for fluency and flexibility.

The results of the experiment were reported the following day. The mean number of acceptable ideas by Group A was 23 and by Group B, 20.4. The mean originality score for A was 6.2 and for B, 12.2. The mean flexibility score for each group was 8.5. It was found that the difference for fluency could be accepted at about the 25 per cent level; the difference in originality could be accepted at the .1 per cent level; and, of course, there was no difference in flexibility. This gave the High Achievers understandable examples of differences in the degree of confidence which can be expected in some experimental studies. It also illustrated the principle that thinking is likely to be more original if originality is rewarded.

CONCEPTS OF MEASUREMENT AND QUANTIFICATION

The homework assignment for the second day of the course was to write an imaginative story on their choice of ten titles involving persons or animals with divergent characteristics, such as:

The Duck That Won't Quack
The Flying Monkey
The Boy Who Wants to Be a Nurse
The Girl Who Wants to Be an Engineer

The stories were used as the basis for developing concepts of measurement and quantification. It was seen rather quickly that the number of words, number of sentences, and the like could be counted and the numbers used to describe the stories. They computed the percentage who wrote on each topic. They also noted that the number of different kinds of words, different kinds of sentences, different kinds of errors, and the like might also be used to describe the stories. They were then asked how they could quantify such qualities as originality and interest. As a result of the ensuing discussion, they named almost all of the criteria for assessing originality and interest developed by Yamamoto and Torrance.

AN EXPERIMENT OF THEIR OWN

At the end of the third session, instructions were given for conducting an experiment of their own. First, there was a pilot study in which the High Achievers served as subjects. They were given the Horse-Trading Problem which is as follows:

> You bought a horse for $60 and sold it for $70. You then decided that you wanted the horse back. You had to pay $80 for it this time, but you sold it again for $90. How did you come out in all this trading? Did you make or lose any money? How much did you gain or lose, if any?

Nineteen of the children said that they would make $10; 14 said that they would break even; 6 said that they would make $20. None said there would be a $30 profit, until the experimenter asked those who had not held up their hands if they had gotten $30—then some of those who had reported other numbers shifted to $30.

They were then given instructions for carrying out their own experiment on the following day. The class was divided into research teams of four or five members each. Each team was to conduct an experiment in a specified classroom in the third, fourth, and fifth grades. On the last day of the course the teams were to present their findings in research reports.

The experiment which the young researchers carried out involved presenting the Horse-Trading Problem to ten of the third, fourth, and fifth grade classes of the school and introducing an experimental variable which could be measured. In order to find out how their schoolmates might be influenced in arriving at their answers, each team collected cards upon which the subjects had written their answers and then "accidentally" dropped a slip of paper which had "30" written on it as they left the room. Later on, when the High Achievers returned to the rooms, they gave the subjects a chance to change their answers before the correct answer was revealed. The subjects were given a dittoed sheet which afforded them an opportunity to change their answers and also asked them if they had discussed the problem with anyone and who had influenced them if they had changed their answers.

When the teams came together they converted some of their findings into percentages and were taught how to use a nomograph to determine whether or not the differences in percentages were significant. The symbols for level of confidence bothered them because they had not been prepared for this, but with this out of the way they were able to determine level of confidence in their own experiments between percentage correct on the first trial and percentage correct on the second trial.

On the last day, the reports came in. In reporting, the High Achievers followed this outline:

1. *Purpose of the Study:* Why did you perform the experiment? What was it that you didn't know? What were you trying to find out?
2. *Procedure:* What was done? Who performed the experiment? What materials did they use? Who were the subjects? How many were there? What did the experimenter do and what did the subjects do? These things should be described in such a way that someone else could perform the experiment and obtain the same results.
3. *Results:* What happened? What did you find out? This should include a presentation of your data in tables, charts, or the like.
4. *Discussion:* What does it mean? Try to explain your results and their meaning.
5. *Conclusions:* What does it all add up to? What can you be reasonably certain about? How certain can you be?

Many of the reports were well organized and were written in good form. A number showed real ingenuity.

A REVIEW AND A TEST

Following a brief review in which an attempt was made to consolidate the gains in knowledge which had taken place during the week, the pupils took a test which dealt with the basic ideas of the course. Judging from their responses, the High Achievers attained a good grasp of the fundamental concepts of educational research.

SOME BY-PRODUCTS

One important by-product from teaching gifted elementary pupils research concepts and skills is the contribution which experiments conducted in such a setting make to the development and testing of new hypotheses, both about the education of gifted children and of others. The historical research project has provided the authors with a number of ideas which had never before occurred to them concerning the development of children and the ways in which parents perceive this development. The experiment conducted by the children helped us to plot in a new way the increase in peer influence and the decline in adult influence in searching for solutions to problems. These and other findings and leads will be presented in other reports.

REFERENCES

GILBERTSON, S. *The second report on the high achievers room.* Bloomington, Minn.: Bloomington Public Schools, 1960.

35

Effects of Structuring on Productiveness of Children's Thinking

Charlotte Antoinette Crabtree

This brief abstract of Crabtree's dissertation illustrates still another link in the investigation of concept attainment in school and the role of interaction between teacher behavior and pupil learning. Crabtree's study is

---

From *Dissertation Abstracts,* 1962, *23:* 161.

also related to those of Torrance (34) and Guilford (3). This abstract is presented to indicate how dissertations contribute to knowledge and how they are reported.

This study was developed to explore some effects of two types of structure upon children's thinking in the years of early childhood. Structure was defined as the organization imposed upon the learning situation. A theory of teacher and environmental structuring was formulated which predicted that a program of child-developed structure, as compared to one of teacher-predetermined structure, would predispose children to greater divergence in thinking, and would elicit from children more "highly constructive" response patterns.

The subjects of this study were twenty-four second grade children. Children were pretested for creativity of thinking on the Torrance Product Improvement Task, paired, and randomly assigned to one of two groups. Both groups received, in alternate sequence, both experimental programs.

Two experimental teaching programs were developed, differentiated by the degree of structure the teacher imposed. t tests of the significance of differences between observed teacher behaviors under the two programs established a high level of confidence that the programs were consistently differentiated over the six weeks of this study ($P < .01$).

In Program B, the Program of Predetermined Structure, Teacher Directive behaviors instituted rigorous limitations upon the self-direction of children's thinking. Children's responses to the circumscribed discussions the teacher initiated were consistently evaluated for their conceptual accuracy against norms of the adult world.

In Program A, the Program of Emergent Structure, Teacher Integrative behaviors opened to children opportunities to explore ideas initiated by them. Children's interests furnished cues for the sequence and pacing of these discussions. Subject matter and teaching aids were introduced as children's cues established their relevance.

These discussion periods of Programs A and B both centered in a study of harbor and airport activities. Following these discussions, children under both programs received equal opportunity to establish, through dramatic play, self-initiated patterns of harbor and airport activities. The teacher at this time assumed, under both programs, the role of a non-participating onlooker. The play environment in Program B was prior arranged with high realism materials, organized to invite certain patterns of response. The play environment in Program A was unarranged, consisted of ambiguous materials, and afforded fewer structured cues for play.

Scales were developed for the measurement of children's thinking and involvement in play. By point time sampling and anecdotal recording, independent teams of observers obtained both quantitative and descriptive data over the six weeks of this study. Inter-observer agreement in processing data into the behavior categories of the scales ranged from 79 to 94 per cent. Tests of significance were computed largely through analysis of variance. Cochran's Q formula was used where highly skewed data required non-parametric statistical treatment. Some final within-groups analyses were tested by chi square.

Divergent thinking, characterized by originality and flexibility of response, occurred more abundantly in Program A ($P < .001$). Convergent thinking, defined as conceptually accurate, consensually validated responses, appropriate to the situation, was more fully elicited under Program B ($P < .001$). Despite no significant differences in their predispositions to creativity, boys expressed more divergent thinking under Program B than did girls ($P < .10$). No significant differences were found between programs in children's involvement in the initial minutes of the play periods. Program A elicited more extensive and highly constructive play sequences than did Program B ($P < .001$). Such responses, however, did not increase within Program A, as hypothesized, but instead showed some drop-off over time. Implications were explored regarding sequence in a curriculum planned to support both divergent and convergent thinking; and questions were raised as to how effectively teachers can mediate the alternate cues found here to predispose and reward both styles of response.

36

Relationship of Self-Perceptions of Early Primary Children to Achievement in Reading

Mary W. Lamy

Lamy's dissertation abstract, like Crabtree's, is too brief to give all the information desired, but her findings do indicate that self-perception in kindergarten is predictive of reading achievement in first grade. This dissertation is part of a longitudinal study by Combs and Soper into the relationship between perception and behavior in the first few years of school.

This research was designed to test an aspect of phenomenological psychology applied to educational practice. Phenomenological psychology postulates that behavior and learning are functions of a person's perceptions of himself and of his world.

This study was designed to explore the role of children's perceptions of themselves and their world in the process of learning to read. Specifically this study was designed to test the following hypotheses:

Hypothesis 1: There is a positive relationship between children's perceptions of themselves and of their world while in *kindergarten* and their subsequent achievement in reading in the first grade.

Hypothesis 2: There is a positive relationship between children's perceptions of themselves and of their world while in the *first grade* and their achievement in reading in the first grade.

Hypothesis 3: A combination of intelligence test scores and perception ratings of children (a) in kindergarten, and (b) in the first grade will have a higher relationship with their first-grade reading achievement than will either set of scores separately.

The *population* for this research consisted of 52 first-grade students, 30 boys and 22 girls, who were enrolled in the University of Florida Laboratory School.

Perception data for these children during both the kindergarten and first grade were available from the research by Combs and Soper, "The Relationship of Child Perceptions to Achievement and Behavior in the Early School Years." The perception data were obtained from inferences made by trained observers working individually with each of the 52 children during kindergarten and first grade. Observations, interviews, and a series of specially devised projective tests were used to provide experiences with each child from which inferences could be made regarding the child's perceptions of (1) his personal adequacy, and (2) his adequacy in dealing with various aspects of his world.

Ten measures of perception for each year provided the basic perception variables which were related to the children's reading achievement and intelligence.

Reading achievement for these children was measured by (1) the California Reading Test, and (2) the first-grade teachers' ratings of reading achievement.

Intelligence was prorated from eight subtests from the Wechsler Intelligence Scale for Children.

Data-Handling Procedures. Relationships between the children's first-grade reading achievement and their perceptions of themselves (1) in kindergarten, and (2) in first grade, were found by Pearson product-moment correlations. Multiple correlations were found between the first-grade reading achievement and a combination of the children's perceptions and intelligence test scores for each year of the study.

Results and Conclusions. The findings of this research have confirmed all three hypotheses regarding the relationships between reading achievement in the first grade and children's perceptions of themselves and of their world.

The findings thus appear to corroborate the general theoretical position concerning the importance of the perceptual point of view. Evidence suggests that a child's self-concept, and the perceptions he holds of himself in relationship to various aspects of his world are not only related to, but may be causal factors in his subsequent reading achievement.

The results of this study indicate that study of a child's perceptions give as good a prediction of reading achievement as do intelligence scores. When these two factors are combined, the predictive power is even greater.

These findings support educational practices which take into account each child's perceptions, and call for activities designed in terms of the meanings they have for the child. Experiences need be sought which provide in the school setting for the development within each child of positive perceptions of himself and of his world.

Unpublished doctoral dissertation, University of Florida, 1962 (2 page abstract)

37

Person Perception in Children

Marian Radke Yarrow and John D. Campbell

The problem of perception exists not only in adult-child relationships but also in peer relationships. It is a crucial concept in social psychology. In this investigation, sponsored by the National Institute of Mental Health, Yarrow and Campbell explore how children view each other and what role the child's own feelings play in his perceptions of his peers. The authors also examine the relationship between peer perception and adult observation of behavior.

How do people "perceive" one another? This question continues to invite research of many kinds. In many of these studies, behavioral responses are taken as perceptual indicators, but the meaning of the other person and the cognitive processes involved in achieving such meaning are left an inference. A needed extension of the work in this area is an inquiry into the ways in which other persons are experienced, as represented by the perceiver's description or conceptualization of the "other." In what dimensions is the other experienced; how uniquely and complexly is he comprehended? How do constant tendencies of the perceiver bring a screen into the situation? How do qualities of the particular other and the particular situation affect perception?

In research on person perception, investigators have proceeded along several different lines. One approach has emphasized *judgmental* aspects of person perception and has focused on factors correlated with an individual's empathic ability, insight, or accuracy in appraising the responses of others. Common to this research is a focus on a discrete judgment, an appraisal of the person perceived along specified dimensions preselected by the investigator, for example, prediction of a person's responses to a picture frustration test, a technique employed by Bieri (1953); prediction of the stimulus person's self description by an adjectival *Q*-sort, used by Baker and Block (1957); or prediction of another's sociometric choices, in work by Tagiuri, Kogan and Bruner (1955).

Another approach less commonly employed was pointed to by MacLeod (1945) when he emphasized the need to ask the phenomenological question "What is there?" More recently Hastorf, Richardson, and Dornbusch (1958), among others, have echoed this view in the statement: "Researchers should make more of an attempt to study the perceptual categories that are actually employed by, and thus relevant to, the perceiver under consideration" (p. 55). Such an approach is basically taken in the work reported in this paper.

This study uses a meaningful real-life social situation as a setting for examining the characteristics of children's perceptions of one another as they occur in a behavioral context. The basic aims of the research are: (a) to examine the dimensions in terms of which children conceptualize and organize their impressions of others; (b) to explore ways in which characteristics of the perceiver and of the perceptual object contribute to interpersonal perceptions; and (c) to assess effects of continued contact (the transition from stranger to acquaintance) on children's cognitive appraisals.

The setting and design of the research utilized the essentially natural laboratory of children's summer camps. The 267 boys and girls, white and Negro, who served as subjects ranged in age from eight to 13. They came from low-income families in the greater Washington area. The children were placed in cabin groups of eight children per cabin. The cabins were homogeneous in age and sex. In half the cabins, groups were racially homogeneous; in half, biracial.[1] The groups were composed of children who were initially strangers to one another. The setting was isolated from familiar surroundings of family and friends. Children of the cabin were, therefore, the child's primary social group for two weeks of camp.

From *Merrill-Palmer Quarterly of Behavior and Development*, 1963, *9:* 57–72.

[1] The role of race in influencing interpersonal perception has not been reported in the present paper, since it has been examined elsewhere by the investigators (1958).

The data on children's perceptions derive from interviews with each child at the beginning and at the end of the two weeks of camp. In the interviews, individual snapshots of all the cabin members were spread out in front of the child, and he was asked to choose one child whom he felt he knew most about. He was encouraged to "tell all about him, as if you were telling a friend back home." The questions were designed to learn what was most salient to the child (whom he selected, and what dimensions he used for description) and the quality of his interpretations. From sociometric questions, each child's ranking of his cabin mates on their desirability as friends was obtained.

For 20 of the 32 cabins studied, systematic behavior observations were made using a sampling of periods and locations in the camp day, on days at the beginning, middle, and end of camp. The observer made detailed running accounts of the behavior of each child in the cabin. This was done by observing for five minutes, followed by five minutes for additional writing-up of observations. In each five-minute record, the observer was required to account for each child present. There were 1,010 such observational records, representing slightly more than four hours in the life of each group studied.

CHARACTERISTICS OF CHILDREN'S PERCEPTIONS OF PERSONS

In order to present the data on children's perceptions of persons it will be necessary first to describe ways in which their "perceptual" reports have been analyzed. Descriptions were divided into psychological units, a unit being defined as a discrete action, or a single characteristic or evaluation. For example, the following response is broken up into twelve units:

> He is nice—and plays all right.—but he likes to fight people.—Says he can beat anybody in the cabin.—He likes to tell jokes a lot. He makes you laugh.—O.K.—but he likes to fight other kids.—Acts all right—except when gets mad, then he wants to punch on people.—Most of the time he is on his bed telling jokes.—He says he don't like to play baseball but he likes to play horseshoes.—That's all I know.

A report was scored on: (a) number of psychological units present, (b) content of the units, and (c) nature of interpretation in the total description.

Content of Perceptual Report

The desire to examine content by using a framework that did not deviate too greatly from the children's phenomenal reports guided the development of a classification scheme. The one evolved has antecedents in the conceptualizations of interaction formulated by Coffey and his associates (Freedman, Leary, Ossorio & Coffey, 1951), but the categories (Table 1) were tailored to the content in the children's reports.

When asked to tell all they knew about another, children showed two very strong tendencies: (a) broad positive or negative judgments were given by 85 per cent, and (b) these judgments were elaborated primarily in terms of peer actions having direct interpersonal consequences. In contrast, noninteractive qualities appeared in only one-half of the reports, as relatively minor fractions of the descriptions. Physical appearance was the most frequently reported noninteractive aspect.

Behaviors described with greatest frequency were "sociable" in nature (Table 1). They dealt with the "business" of camp—play and conversation of the children—described in personal and "feeling" terms—who was friendly, who shared with whom, how a child felt about another, the emotional consequences of play. One may ask whether these positive dimensions reflect strong affiliative needs in a new situation with new companions. Reports of direct aggression, domination, rebellion and nonconformity, which also appeared with considerable frequency, might have parallel immediate interpersonal relevance, though in the negative sense. The appearance of descriptions of fearfulness, dependence, indirect aggressive maneuvers, effective leadership behavior, documents the breadth of children's awareness of interpersonal relations, though the latter categories, which may not be so readily apprehended, occurred relatively infrequently.

Individual children contributed differently to the distribution of characteristics listed in Table 1. On the average, children reported 11 units of description, some reported as many as 27, others a single unit. When each child's description of another was summarized in terms of four major categories of interaction—aggressive, affiliative, assertive and submissive—forty-six per cent of the children concentrated their descriptions entirely in one of these broad dimensions. Thirty-one per cent were able to see opposite or contradictory features in the other child (i.e., both affectionate and hostile behavior, dominating and passive-dependent actions). Only two per cent of the children described the other child in terms of all four categories. The remaining children, in using more than one dimension of content, did not present both ends of any given continuum.

Children's perceptions on the two interviews (in the first or second day at camp and at the end of camp) demonstrated noteworthy stability; content areas appeared in relatively the same frequency positions on the two interviews. Within this consistency, some shifts occurred; emphasis on interaction rather than noninteraction (proportion of interaction units in each description) became even more pronounced. Among children mentioning noninteractional factors in either interview, 56 per cent showed lessened emphasis and 35 per cent increased emphasis. This shift over time was statistically significant. Decline in noninteraction was primarily accounted for by the lesser emphasis on a child's appearance. In descriptions of interaction, reports of dominating and aggressive, rebellious and nonconforming behavior increased and reports of affiliative behavior decreased on the second interview.

Comparisons of the individual child's reports revealed intra-individual consistency. A child's descriptions on the second interview were by no means repetitions of his first interview; however, if a child used a given dimension of personality on his initial description, he was more likely to report it on the second occasion than was a child who did not mention it initially. This self-consistency occurred regardless of whether the same or a different child was being described on the two occasions. Such intra-individual thematic

Table 1: *Content of children's descriptions of others*

Content Category	Illustrative Examples	Per cent of children reporting in interview	
		I	*II*
Interactional content[a]			
(1) domination, direct aggression	"picks fights," "bosses"	31%[b]	49%
(2) indirect aggression	"made fun of me," "chases kids"	15	11
(3) rebellion, non-conformity	"was fresh," "is stubborn"	21	30
1 and/or 2 and/or 3		45%	62%
(4) fearfulness, submission	"scared of everyone," "homesick"	14	16
(5) mild dependence	"asks me for things," "asked for help"	8	6
4 and/or 5		20	21
(6) conformity	"doesn't say bad things," "says 'excuse me' "	50	55
(7) verbal sociability	"talks with us," "laughs at my jokes"	52	47
(8) physical play	"plays all the time," "played games together"	61	55
7 and/or 8		80	74
(9) affiliation	"friendly," "likes to hang around you"	56	47
(10) affection	"likes me," "chose me for a buddy"	32	23
9 and/or 10		69	56
(11) nurturance	"shares," "is a kind girl," "comforted me"	44	51
(12) assertive leadership	"tells me what to do"	18	18
(13) solitary behavior	"sits and watches others," "is quiet"	32	31
Non-interactional content			
(14) abilities, skills, disabilities	"intelligent," "not a good swimmer"	14	15
(15) appearance, features	"has curly hair," "about my size"	35	24
(16) name, age	"got to know his name," "she's ten"	6	6
(17) non-camp contextual factors	"has three brothers," "was a Boy Scout"	10	12
14 and/or 15 and/or 16 and/or 17		51%	42%
Number of children		(247)[c]	(247)

[a] *In a check of reliability, two coders independently classified 77 per cent of a sample of 185 units in the identical detailed categories of content*

[b] *Percentage indicates proportion of children whose description of another child has any content classified in the designated category. Percentages add to more than 100.*

[c] *Not included in this tabulation are those children who, instead of describing another child, described their cabin counselor.*

Table 2: *Complexity of perceptual organization*

Rating[a]	Description of Rating	Interview: I	II
7	Interpretive generalization, supported by interweaving of different complicated themes or by complex analyses of "why" of behavior	8%	8%
6		26%	36%
5		11%	8%
4	Superficial generalization supported by some congruent behavior details	21%	23%
3		20%	16%
2		4%	2%
1	Limited specific details without stated or implied interpretive generalization, or vague global generalization	10%	7%
	(number of subjects)	(261)	(256)

[a] *In a reliability check, two coders independently evaluated 49 children's descriptions and identically classified 81 per cent of these.*

emphasis suggests that each child develops a perceptual framework that for him has general applicability, a possibility that needs testing in further research.

Quality and Complexity of Perceptual Report

Consider these two descriptions of the same child:

> I have been with this guy name of Mark. He always talks back to the chief, and gets me to tie his shoes. I wake him up in the morning . . . He is a funny guy, likes to play around a lot. I get along real good. He acts okay . . . Every five minutes he has to be told something. He likes to act simple, too.

> He is simple. When he talks, he makes me laugh, and he always talks back to the chief (counselor). Yesterday he said that "camp, it looks pretty good, and we look like jackasses." He likes for you to have manners. He likes to play with people. He eats a lot. He always wants food passed around the table. Don't want anybody to put their arms across his plate . . . I think he could get along with others all right. He gets along with me, but everybody picks on him. I like to help out people who everybody pick on, something is wrong with them . . . Last night the man say "have your father been pounding on you a lot?" I would think that, because something is wrong with him. It might make him worser.

These descriptions have used similar categories of content, yet they differ markedly in how they have organized and synthesized the "parts," and in the extent to which they have attempted interpretation of motives or causes of behavior. The description which arrives at a picture-of-a-person or a "theory" of personality may be regarded as more complex than the description which is an atomistic collection of discrete variables.

A seven-point scale of complexity of organization was used to measure the extent to which children gave systematically organized descriptions. Descriptions rated at the top of the scale painted well-integrated personality portraits involving implicit and explicit inferences about behavior. Those in the intermediate range were typified by more superficial generalizations supported by congruent behavioral detail. At the lower end of the scale of organization were the reports that consisted either of vague global generalizations that were inadequately supported (e.g., "He's OK." "She's nice." "He's all right.") or a congeries of specific details without a stated or implied theme (Table 2).

The individual child tended to give descriptions at similar levels of complexity on both interviews, regardless of whether the same or a different peer was described on both interviews. However, there was a significant trend for the group toward more highly organized impressions on the second interview. The qualitative superiority of the later descriptions may derive from the fact that the child had more opportunity to collect many diverse bits of information about another child and to observe more occasions of given kinds of behavior in the child which might aid in the synthesizing process.

From the examination of children's perceptions, several conclusions have emerged: (a) Children's descriptions of others were dominated by social relevance; the emphasis on interaction and the very common use of evaluative appraisals gave evidence of this. (b) There was stability over time both in content and quality of the individual child's perceptual report, whether reporting on the same or a different child. (c) Within this stability, there were certain systematic changes over time toward increased emphasis on interaction and more fully organized descriptions.

There were marked individual variations in children's reports. The next portions of this paper attempt to account for some of these variations by examining certain characteristics of the perceiving child and certain characteristics of the child perceived.

CHARACTERISTICS OF THE PERCEIVING CHILD AND HIS PERCEPTION OF PERSONS

In looking for links between the perceiving child and his appraisal of others, variables of age, sex, and the perceiver's behavior have been considered.

Age and Sex

One would anticipate that the child's conceptions of his human world change as greatly with increasing maturity and experience as do his conceptions of the physical world. There have been very few studies, however, concerning the development of the child's sensitivity to phenomena of human characteristics and behavior. . . . The present study deals with only a limited cross-section spanning middle childhood (eight to 13 years). By this time most children have had experience with a fairly wide range of personalities in the family, neighborhood and school life; also, the differential roles of boys and girls in society are certainly well defined. To what extent are age and sex differences related to variability in children's person perceptions?

Age related trends in categories used in describing others did not appear except in reports of aggression and domination, which tended to be more frequent among the older children (Table 3). There was, however, a sex-linked pattern of differences, consistent with the different experiences of boys and girls, and with adult values for boys and girls. Thus, distinctly nurturing behaviors were emphasized more by girls than boys (especially at the end of camp). The boys, on the other hand, were more attentive to acting-out nonconformities and withdrawn behavior. Attending to the conforming actions in the other child ("She's polite." "He doesn't say bad things.") was also more frequent in boys than girls.

When perceptual complexity is considered, there were no consistent sex differences, but age trends appear. Older children tended to give more complex perceptual reports. A relevant, but unavailable, measure in the present study is the child's intelligence, which might bear a significant relation to his characteristic ways of perceiving others. The relatively small developmental trends are surprising for functions as complex as the perceiving of persons and interpersonal relations. This may be a reflection of the early development of sensitivities to interpersonal relations. Also, more specific differences by age may have been undetected by the analysis (e.g., concrete *details* of aggressive or nurturant behavior might be quite different at different age levels, though the generic categories would be the same). A more detailed developmental study is needed.

Personality as Reflected in Behavior Patterns

The personality of the perceiving child, as expressed through his actions, was also examined in relation to cognitive responses. Behavioral records on each child were divided into units of action. Each unit was classed in one of four broad categories: (a) friendly, sociable, and nurturing behavior (49 per cent of all units), (b) aggressive, rebellious, and disruptive actions (20 per cent), (c) assertive and influencing behavior (21 per cent), and (d) submissive, dependent,

Table 3: *Content of description related to age and sex of reporting child*

	INTERVIEW I						INTERVIEW II					
Sex:	Girls			Boys			Girls			Boys		
Age[a]:	8-9	10-11	12-13	8-9	10-11	12-13	8-9	10-11	12-13	8-9	10-11	12-13
Domination, aggression, rebellion[b]	22%	30%	39%	48%	64%	65%	50%	47%	53%	68%	75%	82%
Fearfulness, submission, dependence	24	19	17	21	21	17	29	17	26	15	26	19
Conformity[b]	34	42	39	56	70	48	44	50	42	62	60	70
Sociability, play	80	92	61	73	83	76	74	77	63	74	74	82
Affiliation, affection	71	75	72	60	74	62	62	57	58	45	64	52
Nurturance[c]	39	55	50	46	36	41	68	65	63	34	34	56
Assertive, leadership	15	11	17	19	19	31	18	13	16	19	25	19
Solitary behavior[c]	15	28	11	33	45	52	29	23	32	40	32	40
Non-interactional content[c]	44	36	56	44	72	59	48	38	32	34	54	48
(Number of subjects)	(41)	(53)	(18)	(52)	(53)	(29)	(34)	(60)	(19)	(53)	(53)	(27)

[a] *No statistically significant differences between age levels were detected.*

[b] *Differences between girls and boys are statistically significant on both interviews. Probability of occurrence of* x^2 *values equal to or larger than those obtained is less than .05.*

[c] *Differences between girls and boys are significant at the .05 level on one interview only.*

fearful behavior (10 per cent). Two coders independently classified 89 per cent of 298 units in identical fashion. Each child was scored in terms of the relative frequency with which he initiated these interactions. Combination of the three most frequently occurring kinds of interaction gave behavioral types defined as follows: "active" children scored in the upper third of their cabin group on friendly as well as aggressive interaction, with assertion ranging from high to average; "withdrawn" children rated in the bottom third on at least two of these three dimensions, and high on none; "aggressive" children rated high on aggressive, disruptive, and rebellious action, *not* high on friendliness, and varied from high to low on assertion scores; and "friendly" children ranked high on sociable, affiliative and nurturing interaction and average or below on both aggression and assertion. Analyses by age and sex revealed no systematic differences in the frequency of these behavior types among boys and girls in the three age groups.

For the analyses that follow, active children were contrasted with withdrawn children on the nature of their perceptions of peers; friendly were contrasted with hostile. Do children who in these respects behave differently give differing cognitive responses? Differences would be predicted either because children of dissimilar personalities may have sensitivities for particular interpersonal phenomena, or because "perceiving" children manifesting contrasting behavior patterns may create differing interpersonal situations. In part the expectation of differences is upheld. Children differing in behavior showed significant differences in complexity of perceptions, but they showed no stable differences in the categories of content used in describing others.

Active and friendly children gave more complex descriptions of others than did withdrawn or hostile children. (Table 4.) Active children made explicit inferences to a greater extent than did the withdrawn (measured only at the end of the session). In comparing friendly with hostile children, there was a consistent, but nonsignificant, tendency on both interviews for the friendly children to describe others in a more organized fashion, and they were significantly more likely than the hostile children to make inferential statements.

These comparisons suggest that the active participant brings his awareness of others into sharper judgmental focus than does the child who remains on the sidelines. Intentional or not, some degree of manipulation of interaction seems to facilitate obtaining a clear picture—accurate or inaccurate—of the social environment. That descriptions by the more sociable and affiliative children were generally more complex than those of the withdrawn children might be accounted for on similar grounds. That descriptions given by the friendly children are not strikingly superior in quality to those given by the hostile aggressive children might suggest that the general equivalence in their amounts of contact with peers tended to outweigh any possible influences of differences in the affective flavor of their behavior.

Table 4: *Complexity of perceptual reports related to selected behavior types*

	Active	Withdrawn	Friendly	Hostile
Median organization rating, Interview I	4.33[a]	3.64	4.50	4.12
Median organization rating, Interview II	5.67[a]	4.33	5.25	4.22
Per cent using inferences	59%[b]	23%	53%[b]	24%
(number of subjects)	(22)	(38)	(19)	(32)

[a] *Difference between "active" and "withdrawn" approaches significance. Probability of occurrence of Mann-Whitney U-value equal to or smaller than those obtained is less than .10.*

[b] *"Active" differ significantly from "withdrawn"; "friendly" differ from "hostile." Probability of occurrence of chi-square values equal to or larger than those obtained is less than .05.*

BEHAVIOR OF THE PERCEIVED CHILD AND OTHERS' DESCRIPTIONS OF HIM

The portrait of a child reported by his peers would not be expected to reflect his behavior with his peers with complete fidelity, but would be expected to show some correspondence. It was possible in this study to examine the closeness or divergence of these two pictures by comparing peer descriptions with records of behavior on each child made by trained observers, covering four hours of interaction. Descriptions and observations of behavior were both reduced to the four broad classes of action—friendly, aggressive, assertive, and submissive. Presence of a given class of action in a peer's descriptions of a child was compared with relative frequency with which the child initiated such action in his group. In this comparison, peer image and recorded behavior showed a remarkable absence of correspondence. In other words, in the child's synthesized picture of another child, the relevance or salience of a characteristic does not appear to be determined by the sheer frequency of its display in the behavior by the perceived child. This lack of correspondence is open to several possible interpretations. In part, it might be a function of the way in which the data were gathered. Thus, the "sample" of behavior available to the adult observer might not adequately represent the "universe" of peer behavior available to the perceiving child. It is possible, too, that the perceiving child selects from the total behavior of the object child only the direct behavioral interchange between himself and the perceived child, whereas the adult observer is attuned to interactions occurring among all the cabin group. This latter explanation is unlikely—because the reporting child frequently referred to the perceived child's relations with others.

While methodological factors may account for some of the discrepancies between the adult's and the child's "conclusions" about the nature of another child, these discrepancies might also derive from "real" differences in their selection from and interpretation of the same raw materials of children's behavior which are important in the understanding of the perceptions of children and which have importance also as a methodological issue in research with children.

Adults and children, while observing the same actions, might attribute quite different meaning to these actions. The values of scientist-observers and those of children—in this instance children from lower socio-economic groups—may differ. It would not be surprising, therefore, if concomitant differences were also reflected in cognitive interpretation of action. Thus, the adult-as-scientist may impute meaning to particular behavior differing from the interpretation given it by a participating child. Behavioral data may be reliably coded, but still such reliability does not eliminate the possibility of the differential perspective suggested. If such possible differences in cognitive interpretations are real and considerable ones, then this would clearly suggest the need for caution in the utilization of adult observers' reports of children's behavior as suitable "objective" indices in some child development research.

A second potential source of discrepancy between adults' records of behavior and children's perceptual reports is that of selective screening. Here the data shed some light and partially support a hypothesis of perceptual selectivity on the child's part. Feelings of the perceiver enter into the level of organization of his descriptions about the perceived child and into the content of his descriptions. Affect of the reporting child and complexity of organization of description were related curvilinearly; disliked children were described in a more systematic fashion than liked children, and liked children were described more systematically than children in a more neutral range of friendship. This relationship was much less pronounced after two weeks of interaction. The interesting qualitative superiority with with which disliked children were described may in part reflect the perceiver's personal concern; it may come as the result of intense negative interactions which furnished conspicuous experiences; it may reflect the need to rationalize a negative evaluation.

Content of the reports reflects the affect of the perceiver in the expected manner. Liked children were described in terms that document "good guy" labels; reports of aggressive, fearful and dependent behavior were less often included. Similarly, the least liked children were reported as less affiliative and less nurturant. The extent to which affect is selectively shaping the perception can be seen by considering simultaneously the affect of the reporting child, the content of his report and the adult observer's behavior record of the perceived child. Even when one controls on the amount of recorded friendly behavior of the object child, the more the child is liked by his describer, the more likely is he to be seen as affiliative (Table 5). The amount of reported affiliation decreases systematically, from highest reported on children liked best, to lowest reported on children liked least. In a similar fashion, still controlling on observed friendly behavior of the object child, children most strongly liked are seen as least aggressive; those least well liked, most aggressive. Now controlling on recorded aggressive behavior of the object child, children strongly liked are systematically seen as more affiliative than children less well liked. And those most strongly liked are seen as least aggressive; those least well liked, most aggressive. Affect appears to influence significantly and systematically the child's selection and interpretation of behavioral information from his peers, resulting in peer assessments quite markedly at variance with the assessments arrived at through the perspectives of scientist-observers.

DISCUSSION

We have examined children's perceptions of persons in a new social situation. Their perceptions, as represented by their reports to the investigator, are generally not unrelated congeries of elements but rather syntheses of varied and sometimes disparate characteristics about the other person. Children appraise others in interactional terms, along lines of personal significance. The most highly organized descriptions are given about children who have affective value for the observing child, particularly negative affective value. The sizable variations

Table 5: *Descriptions of aggression and affiliation, related to affective value of perceived child*[a]

		A. Controlling on friendly behavior of perceived child								
Observed behavior:		High			Average			Low		
Affective value of perceived child:		Very high	Moderately high	Average or below	Very high	Moderately high	Average or below	Very high	Moderately high	Average or below
Report of:	Affiliation	87%	65%	50%	72%	60%	50%	84%	58%	30%
	Aggression	30	59	67	34	60	57	32	25	80
	(*n*)	(30)	(17)	(6)	(32)	(5)	(14)	(25)	(12)	(10)
		B. Controlling on aggressive behavior of perceived child								
Report of:	Affiliation	76%	63%	31%	85%	60%	50%	81%	63%	60%
	Aggression	34	88	77	46	30	67	19	38	40
	(*n*)	(29)	(8)	(13)	(26)	(10)	(12)	(32)	(16)	(5)

[a] *Data for this table derive from initial interviews. Interviews conducted at the end of the camp session gave a generally comparable picture.*

found in children's cognitive reports, both in content and level of organization, present a dimension of individual differences about which there is relatively little information and relatively little understanding as to its significance for interpersonal relations. Some of the variations in perceptual reports were accounted for in terms of characteristics of the child reporters, such as age and sex differences and differences in the behavioral qualities of the perceivers. Older children and the more active and friendly children were somewhat more likely to give complex person perceptions. Differences in content of descriptions by boys as contrasted with girls are consistent with the differences in the experiences of childhood for the two sexes.

Momentary and situational factors as well as enduring personality factors undoubtedly account for additional inter-individual variation. The data on the child's tendency to make appraisals in the same dimensions of interpersonal relationships on two occasions of testing, whether or not the same or a different child was being discussed, give one bit of evidence on this point. There is further evidence in the association found between affect of the perceiver and selectivity in his reports on the perceived child.

It would be misleading, however, to point only to differences in children's perceptions. Children's appraisals of their peers in many respects showed shared sensitivities, with repeated thematic emphasis on certain salient relationships, such as sociability, affiliative tendencies, potentially hostile acts. The same themes, with few exceptions, were characteristic of both initial and considered impressions at each age level studied. Thus, the *variations* among children are within this broad common framework. These similar sensitivities (unchanged by age) may possibly be attributable to a common culture of childhood and to similar learning of the significant interpersonal cues, those that are differentially rewarding or punishing in nature.

The particular characteristics of children's person perceptions obtained in this study in unknown ways may derive from the characteristics of the population being studied (children from lower socio-economic classes), and from the particular circumstances under which the testing was carried out. The children were in a new situation in which the other children, the environmental setting, and the adults in charge were all unfamiliar to them. These factors may possibly have heightened children's sensitivities to particular kinds of qualities in other persons. Lack of correspondence between children's reported impressions of another child and adult observers' behavioral records of the child was far greater than one might have anticipated in advance. The children's personal perceptions, fashioned, we assume, from expectations, past experiences and personal needs, resulted in "realities" quite different from the assessments of children based on the observations and codings of the researchers. The coexistence of such discrepant "realities" poses a complicated problem in understanding children's responses to interpersonal stimuli, and one calling for more research. It is a truism to state that person perception is a complex process, yet the evidence of the present study indicates the need for emphasizing this fact.

REFERENCES

BAKER, BELA O., & BLOCK, J. Accuracy of interpersonal prediction as a function of judge and object characteristics. *J. abnorm. soc. Psychol.*, 1957, *54*, 37–43.

BIERI, J. Changes in interpersonal perceptions following social interaction. *J. abnorm. soc. Psychol.*, 1953, *48*, 61–66.

FREEDMAN, M. B., LEARY, T. F., OSSORIO, A. G., & COFFEY, H. S. The interpersonal dimension of personality. *J. Pers.*, 1951, *20*, 143–161.

HASTORF, A. H., RICHARDSON, S. A., & DORNBUSCH, S. M. The problem of relevance in the study of person perception. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior.* Stanford: Stanford University Press, 1958. Pp. 54–62.

MACLEOD, R. B. The phenomenological approach to social psychology. *Psychol. Rev.*, 1945, *54*, 193-210.

TAGIURI, R., KOGAN, N., & BRUNER, J. S. The transparency of interpersonal choice. *Sociometry*, 1955, *18*, 624–635.

YARROW, MARIAN R. (Issue Ed.) Interpersonal dynamics in a desegregation process. *J. soc. Issues*, 1958, *14*.

38

New Patterns of Relationships Between the Sexes Among Preadolescents

Carlfred B. Broderick and Stanley E. Fowler

This short paper by Broderick and Fowler examines scientifically an idea that has been observed by many lay people: that children's relationships with each other, particularly across sex lines, have changed in the last decade. As the authors indicate, there is a growing need for additional study in this field because of the continuing change in these relationships and their impact on our society.

The first purpose of this paper is to document the emergence of new norms for cross-sex interaction among preadolescent children. The second is to offer an interpretation of this new development. The need for new research in this area can be appreciated if one compares descriptions of preadolescent social life written in the 1930's with the findings of a survey published in 1958 by the U.S. Office of Education.

In 1930 Furfey summarized his careful and systematic observations of boy-girl relations at this age as follows:

> ". . . girls are rigorously excluded from participation in masculine activities. The girl, however, does not feel the affront very keenly since she . . . has the same negative attitude toward boys that they have toward her" (Furfey, 1930).

This observation is reinforced by Campbell's 1939 finding that:

> (The boy), ". . . becomes sufficiently conscious of sex so that he does not wish to touch girls or show them any attention except under socially approved conditions such as in games or dancing." (The girl), ". . . begins to be sufficiently conscious of sex so that she will not deliberately touch boys except under conventional circumstances as in games or dancing. She classifies games according to sex—boys play this, girls play that" (Campbell, 1939).

By contrast in a recent national survey, fourth, fifth, and sixth grade teachers gave quite a different picture:

> ". . . in some schools boys and girls do not seem to feel a strong need to separate. In fact, as low as the 4th grade and continuing through the 6th, they frequently ask for activities such as folk dancing and table games *together*, and dating begins in some cases . . . in 5th grade . . . Boys seem to do more personal grooming (some begin in the 4th), carrying a comb and using it, washing hands voluntarily, even occasionally wearing a tie. Girls begin to wear lipstick and nail polish and to dress up their hair. A few children wear 'steady rings,' and some date. . . . In grade 6, . . . 'They show they like each other in a friendly way.' 'They aren't so antagonistic as formerly.' 'They don't pick on each other so much.' 'They get along better.' 'Their social relations are much more mature' " (Lewis, 1958).

It will be seen that although the old pattern of avoidance may still be a potent factor in many groups, new patterns are emerging which promise to revolutionize boy-girl relationships at these ages.

In order to gain more information on these developments, in the spring of 1958 we studied the cross-sex interaction patterns of 264 fifth, sixth, and seventh graders in an elementary school located in a middle class district in a Southern urban community. Although detailed socio-economic data were not available, the school records of parental occupations indicate that a large majority of these children come from homes in the upper lower to upper middle class range. The age range of the entire sample was 9 to 13 years for girls and 9 to 14 years for boys. With relatively few exceptions the 108 fifth graders were 10 or 11, the 79 sixth graders 11 or 12, and the 77 seventh graders 12 or 13 years of age. Although our sample is drawn from a single community, it is representative in that it reflects the trends reported in the previously mentioned national survey.

THE EMERGING PATTERN

Nothing better illustrates the thawing of relationships between the sexes over the past quarter of a century than a comparison of friendship choices at these ages. Studies done in the twenties and thirties report univocally that the percentage of friendship choices extended across the

---

From *Marriage and Family Living*, 1961, *23:* 27–30. Reprinted with permission of the National Council on Family Relations.

Table 1: *Friendship choices by sex of friend chosen and by grade**

Grade	PER CENT			Total Number
	Cross-Sex Choices	Same-Sex Choices	Total	
5th	19.7	80.3	100.0	350
6th	16.0	84.0	100.0	256
7th	14.6	85.4	100.0	253
Total	17.1	82.9	100.0	859

* *Each child had up to four choices. The units in this table are choices.*

Table 2: *Frequency of choosing at least one friend of the opposite sex, by grade**

Grade	PER CENT			Total Number
	One or More Cross-Sex Choices	No Cross-Sex Choices	Total	
5th	51.9	48.1	100.0	108
6th	41.8	58.2	100.0	79
7th	37.7	62.3	100.0	77
Total	44.7	55.3	100.0	264

* *Each child had up to four choices. The units in this table are children.*

Table 3: *Frequency of sweetheart choice, by grade and sex*

Grade	PER CENT				Total Number	
	Have a Sweetheart		No Sweetheart			
	Boys	Girls	Boys	Girls	Boys	Girls
5th	86.8	94.5	13.2	5.5	53	55
6th	86.4	97.1	13.6	2.9	44	35
7th	74.4	86.8	25.6	13.2	39	38
Total	83.1	93.0	16.9	7.0	136	128

sex barrier dropped to near zero in about the third or fourth grade and remained there through about the eighth, after which a slight rise was discernible. (See for example Hsia, 1928; Seagoe, 1933; and Moreno, 1934.) By contrast, Table 1 shows that when the children in our sample were asked whom they liked best of all the children they knew (four choices were permitted) the choices across sex lines ranged from 19.7 per cent in the fifth grade to 14.6 per cent in the seventh. If we look at the percentage of children who chose at least one of four friends across sex lines, (Table 2) we find even more dramatic evidence of the new look in boy-girl relations at this age. From 51.9 per cent of the children in the fifth grade to 37.7 per cent of the children in the seventh grade chose one or more friends of the opposite sex. It would seem from this that although most boys and girls still prefer the companionship of their own sex, many have bridged the gulf between the sexes.

In addition to this increase in cross-sex friendship, there has been an increase in romantic interest among these children. Most probably romance has always been an element for some children of this age. For example, in 1930 Furfey described Max, a 12 year old boy, who openly and ardently declared his affection for his sweetheart. In commenting on the situation, however, Furfey wrote, "Although love affairs are common enough among 12 year olds in fiction, a case like Max's is very rare in real life. This devotion to a girl is uncommon before adolescence and is a distinctly precocious trait" (1930). Campbell, in describing the typical 12 to 13 year old girl, wrote, "She would not admit that a certain boy is attractive to her, though she begins to take a covert interest" (1939). This picture of generally disinterested boys and covertly interested girls contrasts sharply with the situation our study reveals. In our sample the great majority of children in each grade claimed to have a sweetheart, (Table 3) and most of these expected reciprocation (see in Table 4 the responses to the question, "Does Sweetheart like you too?"). Moreover the majority of them apparently do not keep their feelings to themselves (see in Table 4, "Does Sweetheart know?", "Do your friends know?", "Do your folks know?"). That these children in fact do discuss their romances with others is borne out by the ability of their classmates to list the sweetheart pairs in their schoolroom with considerable accuracy, especially in those cases where the relationship was reciprocal or believed by one member to be reciprocal.

Apparently, the question of dating was not investigated during the period which we have used for comparison (1930–1939). In 1949, Hollingshead (1949) found that the most "adventurous" youngsters begin to date at the age of 12, and that among the 13 year olds in Elmtown about 15 per cent of the boys and 20 per cent of the girls had begun to date. Later studies by Smith (1952), Lowrie (1956), and Cameron and Kenkel (1960) indicate the increasing incidence of dating at early ages. The 1958 Office of Education survey previously mentioned indicated that in some schools dating may begin as early as the fourth grade. Our own data illustrate this trend (Table 5), for in the fifth grade (10 and 11 year olds) 45 per cent of the boys and 36 per cent of the girls claim to have had dating experience. By the seventh grade (12 and 13 year olds) nearly 70 per cent of the boys and 53 per cent of the girls claim to have had at least one date.

Table 5 also indicates that some experience with kissing is common at these ages. Unfortunately we know of no comparable data from the earlier period.

Although the data we have seen so far are impressive, some of the most convincing evidence that times have changed comes from another portion of our study. Children were asked to rank the desirability of a companion of the same sex, a companion of the opposite sex, or of no companion at all in three different situations, (eating, taking a walk, going to a movie). They could rank the cross-sex companion as first, second, or third choice. Table 6 shows the results. In the sixth and seventh grades, the majority of boys and girls felt that when taking a walk or going to a movie, the companionship of the opposite sex is to be preferred above either of the alternate arrangements. Both sexes are more conservative when choosing an eating companion, but in the 7th grade the proportion preferring the opposite sex rises to nearly one-half even in this setting.

EXPLANATION OF DATA

In order to account for the changes which we have noted it may be profitable to look at the reasons that are generally put forth to explain the mutual withdrawal of the sexes when it occurs. Blair and Burton (1951) summarize the conclusions of most writers in this field when they list three explanations for cross-sex antagonism. Let us consider each in its turn.

(1) ". . . it is probably partially based on the differences in physical and intellectual development of boys and girls . . . ," referring to

Table 4: *Frequency of "yes" answers to sweetheart questions*

Question	PER CENT*	
	Boys	Girls
Does sweetheart like you too?	52.2	51.3
Does sweetheart know?	68.1	68.9
Do friends know?	63.7	73.1
Do folks know?	62.0	74.8

** Per cents based on 113 boys and 119 girls who made a sweetheart choice.*

Table 5: *Dating and kissing experience, by grade and sex**

Grade	PER CENT					
	Dated		Kissed Opposite Sex		Been Kissed by Opposite Sex	
	Boy	Girl	Boy	Girl	Boy	Girl
5th	45.3	36.4	58.5	32.7	66.0	65.5
6th	52.3	28.6	79.5	34.3	72.7	54.3
7th	69.2	52.6	89.7	44.7	87.2	86.8
Total	54.4	39.1	74.3	36.7	74.3	68.8

** These percentages based on full sample (see Table 3 for marginals).*

Table 6: *Preference for opposite sex companion in three situations, by grade and sex**

Situation	Grade	PER CENT	
		Prefer Opposite Sex Companion	
		Boys	Girls
Eating	5th	35.0	30.9
	6th	34.9	31.4
	7th	46.2	47.4
Walking	5th	49.0	50.9
	6th	69.8	51.4
	7th	71.8	68.4
Movie	5th	58.5	45.5
	6th	68.2	51.4
	7th	74.3	65.7

** Percentages based on full sample (see Table 3 for marginals).*

the fact that at this stage girls are approximately six months more mature physically and mentally, and often a year or more advanced socially. These differences in themselves are considered to be less important as causes of hostility than the differences in interests and activities which they engender. Since these differences are continuing ones, we must assume that they exert an influence toward the separation of boys and girls of the same age today as they did a generation ago. Therefore, we must look elsewhere to find the factors which have given impetus to the changes in boy-girl relations.

(2) " . . . antagonism is partly due to the effort of individuals to identify themselves more closely with their own sex." Through mutual withdrawal each group provides support for the values distinctive of its own sex roles. This is reinforced through the systematic rejection of the competing values of the opposite sex role. Boys would seem to have additional motivation for withdrawal, for at this age it is particularly difficult to maintain the superiority of the masculine role in the face of the double threat presented by the superior status of mothers and female teachers, and the superior academic and social achievements of girls the same age. The traditional masculine tactic seems to be disdain for the feminine values and withdrawal from competition with females into a world of strictly masculine interests. There is time enough to emerge when he may assert his superiority from a stronger position in adolescence. Many of these same elements are involved in Blair and Burton's third explanation of antagonism also.

(3) "Perhaps it is a result of instilling into children the difference between boys' and girls' roles almost before birth." Traditionally boys and girls have had such different behavior and values expected of them that a period of role-training and practice in relative isolation from the opposite sex was a necessary preliminary to their attempt to fulfill these roles later in a heterosexual setting.

These last two explanations for cross-sex antagonism during preadolescence are based on differences in both the *status* and the *content* of traditional male and female roles. There is considerable evidence, however, that in large segments of our society these differences are diminishing. For example the 1955 Purdue Opinion Panel and a study by Landis (1958) both indicate that in the last generation the sex roles have become much more flexible and now overlap in many areas. The contents of the two sets of expectations are becoming more similar as women have achieved many masculine prerogatives and men have begun to share many traditionally feminine responsibilities. As these roles converge and the experiences and values of the two sexes become more similar, cross-sex hostility becomes less appropriate. Rejection of the values of the opposite sex loses much of its purpose when their values are very similar to one's own. Similarly, as the social statuses of the two sexes approach each other many boys seem to feel less need to defend a shaky claim to superiority during this period.

In addition to the negative effect of lessening the need for defensive withdrawal, increasing consensus in important values has been shown to provide a positive impetus to the development of friendly relations between the sexes.

CONCLUSION

We set ourselves the task of documenting the emergence of new norms in the relationships between the sexes among youth 10–13 years of age and to interpret the trend. While old patterns of hostility and withdrawal are not dead, new behaviors and relationships are developing, based on greater understanding and sharing of value orientations. Further research must determine how widespread or unequal these changes may be throughout the nation. Yet if our interpretation of our findings is correct we may expect continued changes in the direction of greater cross-sex interactions and friendships, including romantic attachments at this age. With these may be expected more early cross-sex sexual experimentation, an increase in early marriage rates and greater difficulties in finding clearly defined and generally accepted social sex roles.

REFERENCES

BLAIR, A. W., & BURTON, W. H. *Growth and development of the preadolescent.* New York: Appleton-Century-Crofts, Inc., 1951.

CAMERON, W. J., & KENKEL, W. F. High school dating: A study in variation. *Marriage Fam. Living,* 1960, *22,* 74–76.

CAMPBELL, ELISE HATT. The social-sex development of children. *Genet. Psychol. Monogr.,* 1939, *21,* 461–552.

FURFEY, P. H. *The growing boy*. New York: Macmillan, 1930.

HOLLINGSHEAD, A. B. *Elmtown's youth*. New York: Wiley, 1949.

HSIA, J. C. *A study of sociability of elementary school children*. New York: Teachers' College, Columbia University, 1928.

LANDIS, J. T., & LANDIS, MARY G. *Building a successful marriage*. 3rd ed. Englewood Cliffs, N.J.: Prentice-Hall, 1958.

LEWIS, GERTRUDE M. *Educating children in grades four, five, and six*. Washington, D.C.: U.S. Office of Education, U.S. Department of Health, Education, and Welfare, 1958.

LOWRIE, S. H. Factors involved in the frequency of dating. *Marriage Fam. Living*, 1956, *18*, 46–51.

MORENO, J. L. *Who shall survive*. Washington, D.C.: Nervous and Mental Disease Publishing Co., 1934.

SEAGOE, MAY V. Factors influencing the selection of associates. *J. educ. Res.*, 1933, *27*, 32–40.

SMITH, W. M., JR. Rating and dating: a re-study. *Marriage Fam. Living*, 1952, *14*, 312–317.

COGNITIVE DEVELOPMENT

39

The Development of Scientific Concepts in Children

W. H. King

Basing his research on Piaget's earlier work, King investigates concept development in British children of primary school age. To what extent are scientific concepts acquired with age and on the basis of past experience? And more important, what is the role of instruction and guidance? King tries to give answers based on the results of problems dealing with length, time, volume, weight, night, etc. See Ojemann and Snider (33) for another discussion of the influence of instruction on behavior and development.

INTRODUCTION

Piaget's pioneer work in the field of development of children's concepts of causality has been the starting point for research into the growth of scientific concepts. He found a sequence of seventeen distinct types of causal explanation; each was characteristic of different stages of development and the main groupings made their appearance at four different ages. "The first two stages are precausal, whereas the third, appearing at about seven or eight years is more truly causal. These stages of causal thinking are tied up with the progress from realism to objectivity, from egocentrism to socialization in the child" (Deutsche, 1937). Piaget (1930) believed that the child's explanation of physical phenomena before seven or eight years of age was not in terms of scientific and physical causality and stressed, in the early years of a child's life, his tendency to endow physical objects with human feelings; this animism also developed in four stages. . . .

. . . Johnson and Josey (1931) repeated Piaget's experiments in logical thinking and failed to find evidence of animism, mysticism, etc.; instead, at 6 years of age, they found children were capable of adopting hypotheses and displayed no egocentrism that hindered reasoning. Again, Isaacs (1930) found that mechanical causality appears spontaneously much earlier than the ages suggested by Piaget and suggested that the social and physical background under which Piaget's researches were carried out were unfavourable to its growth and manifestation. Hazlitt (1930) found that children can see relationships at a very early age and that "Piaget's

---

From *British Journal of Educational Psychology*, 1960, *31*: 1–20.

picture of the difference between adult and childish thinking is due to an overvaluation of verbal expressions as a measure of thinking and to an exaggerated idea of the logicality of adult thought."* Oakes (1947), in a detailed research on children's explanations of natural phenomena concluded:

(1) Each subject, regardless of age, mental ability, or grade level, gave explanations of a wide variety of types. All types of answers were given by all age groups.

(2) No evidence was found to corroborate Piaget's interpretation that there is a definite stage in the child's thinking which is characteristic of a given age. The types of answers given by these subjects were influenced more by the nature of the problem, the way the question was worded, the child's background of experience, and his vocabulary than by any other so-called mental structure for a given age.

(3) Although a few responses were enigmatic, the great majority were matter-of-fact, nonmetaphysical; in other words naturalistic.

(4) In general, understanding of essential relationships increases with age.

Oakes's answers to the questions (e.g., What makes clouds move?) were obtained verbally from personal interviews with a large number of children from different environments. Piaget's clinical methods have been criticised on the grounds that he has usually dealt with small groups not drawn from a variety of backgrounds and that he has usually omitted to investigate the socio-economic background which would affect speech development and children's experience and interests. Again the adult-child relationships in an interview situation can determine the nature and content of the response and verbal replies depend on the stage of speech development of the child. In practice, numbers have to be limited. Ideally, research into the nature of children's scientific concepts might well contain a questionnaire type of situation with a follow-up of the answers by interview. The results of the research now reported depended on obtaining information by the first method by seeking the help of the teachers who normally took the children and by making sure that the questions were worded and presented by their teachers in a way which followed normal classroom procedure; in this way it was hoped to minimise the effect on children in the age groups of a test situation. Teachers reported that the questions themselves supplied the motivation and, judging by the discussions that took place afterwards, the co-operation of the children was freely given. A summary of one of these discussions has been added where relevant. Age trends, based on percentages of correct answers for boys and girls of 6 to 11 years were obtained. Some of the results showed clear and unmistakable increase with age; a few disclosed little pattern. In these cases, concepts contained in the questions may not have been understood or were not understood at particular ages. Deutsche gave a justification for this method of scoring. "If the answers given show a progression from low to high scores with increasing age, this would seem to be a partial validation of the method of questioning and obtaining the answers, and of the method of scoring. Unless the answers reflect the approximate level of ability of the subject, this age progress would scarcely be found."

THE INVESTIGATION

Material

A schedule was drawn up containing seventy questions arranged under five headings:

Section A—Length, weight, time, direction.
Section B—Volume and weight.
Section C—Mainly mechanical principles, lever, wheels, etc.
Section D—Living things; seasons, etc.
Section E—Shadows, sections, etc.

These questions were discussed by a panel of primary school teachers and a final form accepted as suitable material for children in junior schools. The results of Section E are not reported here.

Procedures

The schedules were given to the classes over several weeks: they were divided, at the discretion of the teachers, into a suitable number of questions for each session. Questions were not given to children as part of a class test situation, but care was taken to ensure that a normal classroom atmosphere was maintained and normal encouragement given short of any suggestions that a particular form of answer was required.

---

* In his more recent work, with Inhelder, Piaget has realized this weakness and relies more on practical experiment and detailed questioning than on mere verbal expressions.

Six questions in Section A required an estimate of time, length and weight. . . . Answers to all the items were recorded under chronological age and under the headings, where appropriate, of yes, no, don't know. Several answers to items in Section D were written down by children (or recorded by the teacher in the case of younger children).

Subjects

1,811 children in the age range 5 to 12 years from twenty schools took part in this experiment. There were too few children of 5 or 12 and their scores were omitted from the analysis. Absences of children, changes of staff and incomplete returns of all items meant that 1,235 children finally completed the whole of the schedule, most of the absentees occurring at points in Schedule E. Many of the questions were completed by larger numbers; thus, one topic, taken at random from Section D, produced 1,783 replies. In one school, 291 children aged 5 to 11 completed all sections of the schedules, five schools completed the work with children in range 7 to 11; the remaining schools were represented by children in four or fewer of the age groups.

Distributions by age, sex and the means and standard deviations of I.Qs. on the Otis Alpha Test A of the 1,235 children are given in Table 1. The decrease in means with increase in age suggests that teachers probably gave these schedules to their top classes of the age groups 6, 7, 8; on the other hand, the decrease in standard deviation with increase in age shows that this Otis test did not give the older children enough head room. The differences between the means of boys' and girls' scores were significant only at the age levels 6 and 7 years.

Table 1: *Results of the Otis Alpha Test A*

Age	N_B	N_G	x_T	α_T
6	30	34	110	14.6
7	56	81	111	12.6
8	85	107	109	11.6
9	131	156	106	9.3
10	176	174	105	9.9
11	103	102	103	9.0
	581	654	—	—

RESULTS

Section A

Estimation of time: 15 seconds and 45 seconds (Figures 1 and 2). Estimation of time has been found very difficult even for adults. . . . Figures 1 and 2 show that the 6-year-old children tended to under-estimate both times; those aged 8 years to overestimate both times; but throughout the age range there is a steady increase in correct estimations to 67 per cent for the shorter time and 49 per cent for the longer time. Graphs were drawn separately for boys and girls for each time; no significant sex differences were noted and their estimations were combined.

Estimation of distance: Length of school blackboard and length of school building, fence, etc. (about 50 yards). Figures 3 and 4. Most children underestimated the lengths, particularly the longer one. Boys and girls gave the same pattern of responses. At 10 years for boys and a year later for girls, percentages of correct and overestimated replies were equal, from these ages onward percentage correct answers increase at expense of overestimated. For the longer length there was definite evidence of increase in accepted estimations with age, but correct responses were small compared with the underestimates. This may be due to the fact that children have for both measurements to express answers in units with which at this stage they are not too familiar.

Estimation of weight (Figure 5). The children were asked to estimate weight of a milk bottle when full of milk and when empty. The distribution of replies with age were similar for boys and girls and followed similar patterns to those obtained for distance. There was an even more marked tendency to underestimate both weights, and while there was a slight steady increase of correct responses, the final figures reached for boys and girls was only 20 per cent. . . .

Estimation of direction (Figure 6). Two similarly worded questions were given: in the first the children had to choose one of five given directions in which to throw a ball as far as possible, and in the second, to choose one of the same five possible directions at which a child would have to jump if he were to cover the greatest possible distance. Five possible alternatives were offered varying from the horizontal, through angles of 22½°, to the vertical. Theoretically this

Fig. 1: *Estimation of time—15 seconds.*

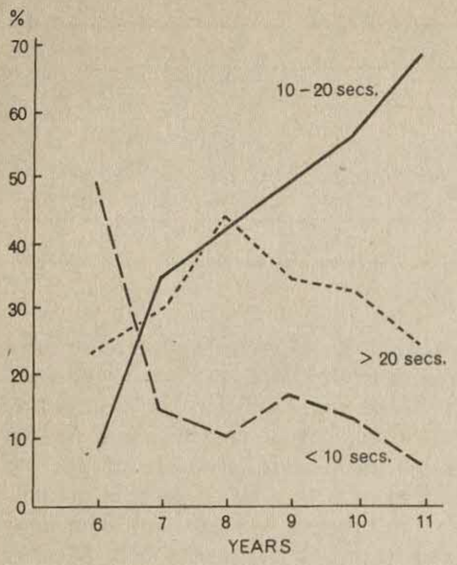


Fig. 2: *Estimation of time—45 seconds.*

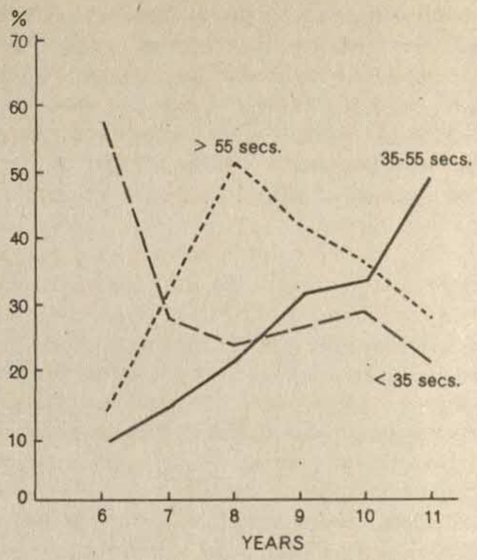


Fig. 3: *Estimation of distance—shorter distance.*

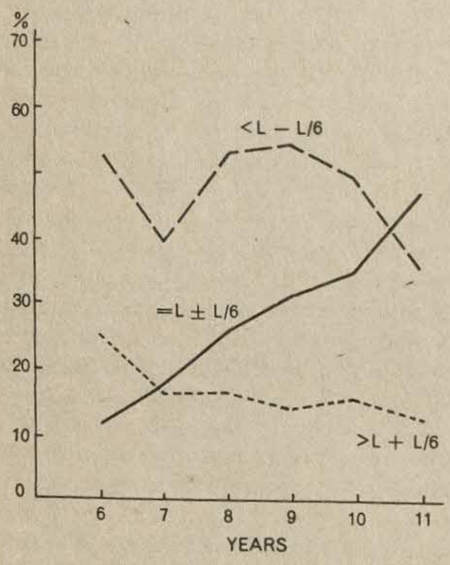
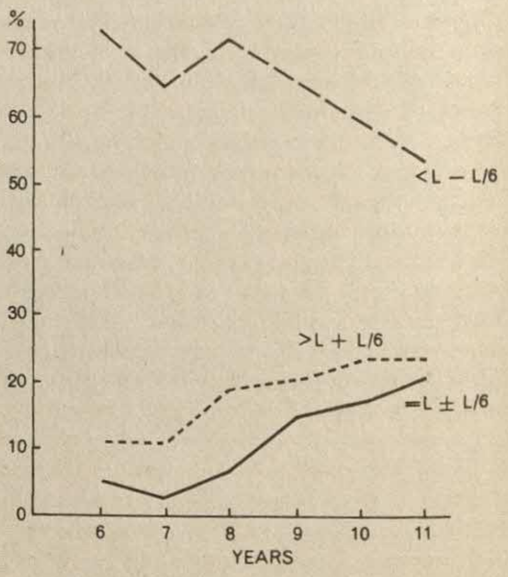


Fig. 4: *Estimation of distance—longer distance.*



is fundamentally the same problem but each question proved difficult to children and to some adults.

The ball problem gave a steady increase in correct responses from 6 years (16 per cent) to 9 years (50 per cent) with much the same responses at 10 and 11 years. 37 per cent of 6-year-old children affirmed that the vertical direction was required, but this percentage dropped smoothly to 10 per cent at 11 years. The pattern of responses for the question on jumping gave no conclusive result. The 22½° angle of projection, with horizontal was popular with the 9-, 10- and 11-year-olds (about 35 per cent), the angle of 67½° consistently unpopular.

At one school discussions were held with the teacher. "It would appear that the large number of children selecting the correct answer (for direction of ball), did so as a compromise between the vertical, which would give height, and the horizontal which would ensure distance. They agreed that, unless you had height, the ball would soon touch the ground (effect of gravity), but unless it was thrown forward, it would not reach far enough."

When a person, not an object, is under consideration, a different method of reasoning is adopted. "A 9-year-old girl stated that she jumped upwards to gain height and then moved forward in the air, the ball she explained could not do this, therefore, in that case, you had to give it height and distance by throwing the ball at an angle of 45°." This is certainly a physical explanation limited by the experience of the child; naive, plausible, illogical by adult standards but certainly not animistic; it is nearer to the type called by Piaget (1929) Mechanical Causality.

Section B

The eight questions, mainly on volume, in this section were presented, in pairs. "When a stone is added to water in a jar, is the level of water higher, lower, or the same as it was before?" There was a steady increase for both boys and girls in correct responses from 54 per cent at 6 years to 89 per cent at 11 years (Figure 7). Piaget found that true explanation did not take place until 10 years, but no explanations were available in this study. When sand was substituted for stone, the responses changed completely; there was no regular increase with age (Figure 8). This question was too difficult and chance rather than sound reasoning determined the form of answers at all ages. The comments from one school showed that experience did not really help the children. "Such experience as the children could call upon when faced with this question was limited to their visits to the seaside. Several children said that the sand 'soaked up' the water; one boy (10 years) explained that sand is not solid and, therefore, would not have to push the water away when it was poured into the jar. Another boy of 11 years asked if it was wet or dry sand. He was told that it was dry, and he gave the answer that the level of the water was unchanged. If it had been wet, he explained, it would depend how wet. Another boy argued that a pond near the school had been filled in with a mixture of earth, clay and sand; the water had not overflowed onto the road."

The next two tests dealt with conservation of volumes. . . . The bases of the four flasks are equal. Water from [the first] flask is emptied into [the second] flask and the probable new levels

Fig. 5: *Weights.*

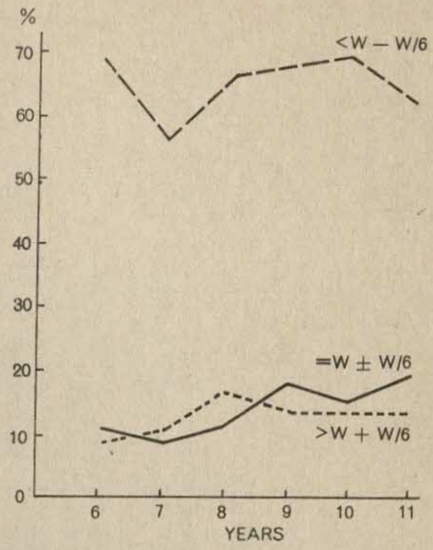


Fig. 6: *Direction.*

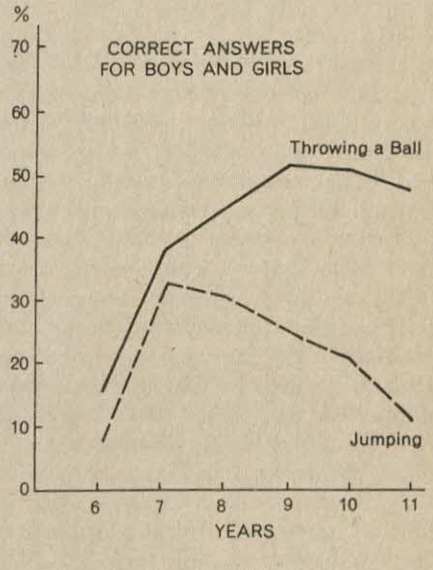
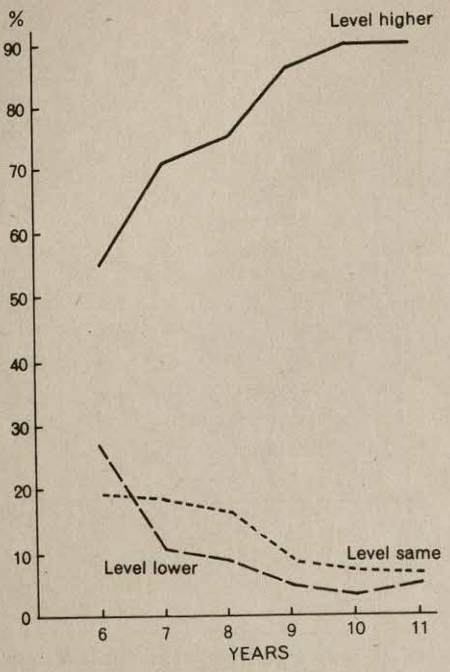


Fig. 7: *Stone added to water.*



of water stated in [the second] flask. The first problem showed a steady increase in correct percentage response with age and the two sexes showed the same trends. Correct answers were given by 45 per cent at 6 years and 90 per cent at 11 years. In the second case there was evidence of some confusion of thought, possibly a carry over from the previous question. More than half of the children ages 6–8 thought that the level now was unchanged, and it was not until 8 years of age that there was a general upward trend which reached the 50 per cent level at 11 years. Boys and girls followed the same pattern but boys obtained higher correct scores at all ages. The difference was 10 per cent at age 11 years.

A jar with vertical sides was filled with water, and the children were then told that all of the water could be poured into any one of five differently shaped containers. They were asked to state which container now held the largest volume of water or whether the volume of water was unchanged. As a corollary they were also asked which containers held the greatest weight of water or would they all hold the same weight. Diagrams of the five containers were drawn on the blackboard and care taken to give clear descriptions; in spite of this the percentage replies to the six answers showed confusion at all ages. However, there was a clear trend as age increased to give correct answers to the question on weight. Girls (24 per cent at 6 years to 41 per cent at 11 years) were superior to boys.

The last two questions in this section concerned two lumps of plasticine shown to be the same in size and weight. One lump was rolled into a cylinder and the children were asked if it was still the same in weight and volume. There was an irregular rise in correct responses and a corresponding decrease in incorrect responses over the age range. At 11 years of age, approximately 80 per cent of the children agreed that the weight remained the same. It is interesting to note the comparatively large percentage of boys (15 per cent at age 6 and 10 per cent at 7 years) who were not prepared to give an answer. A much larger percentage of boys (28 per cent) of this age refused to state whether the plasticine was the same volume or not. Nearly 80 per cent of the 6-year-old girls and 53 per cent of the 6-year-old boys stated that the volume was not the same but after the age of 7 years only a slight majority at each subsequent age gave as their preferred reply that they were not the same volume. From 8 years upwards only a negligible percentage of children (0 to 2 per cent) failed to supply an answer to either of these questions. It was clear that an appreciation of constancy of weight increased with age but that constancy of volume was not understood. It was most unlikely that either of these topics had been dealt with formally in the classroom, but experience gave correct answers to the one, not to the other.

Section C

The first topic in this section dealt with solutions of well-known substances and the different forms of water. Each class was reminded that ice turns into water and washing soda dissolves in water. The children were then asked which of the following substances, tea, sugar, steam, hail, salt, clouds, soap and snow, are able to turn into water. The percentages of boys and girls who gave the correct answers were recorded again under chronological age. Tea, sugar, salt, soap all gave fairly consistent answers; boys and girls, except for about 10 per cent at each age, gave correct answers. The remaining answers on steam, hail, clouds and snow, all showed increasing percentages of correct answers with increase of age. There was no significant sex differences at any age.

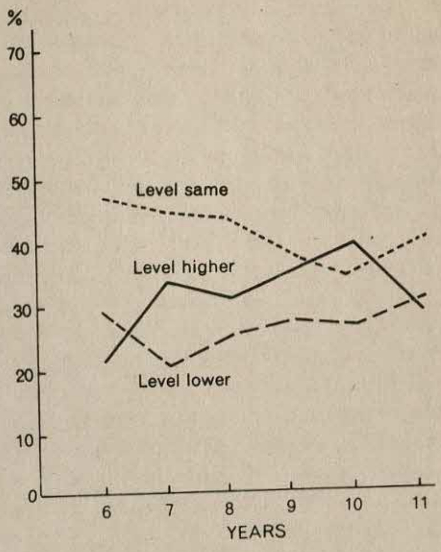
To the question "Does hail turn into water?" 30 per cent at 6 years gave a correct response. This figure rose rapidly to 75 per cent at 9 years and to 78 per cent at 11 years. Seventy-five per cent of the children at 9 years gave the correct reply rising to 87 per cent at 11 years on the change of steam to water. There was also a steady increase in the realization that clouds and water were similarly connected although the final correct answer was given by 71 per cent at 11 years, 42 per cent at 6 years and 40 per cent at 7 years. The relationship of snow with water, quite understandably, was clearer than for the other similar questions. At 6 years 82 per cent gave correct answers quickly rising to 94 per cent at 8 years which percentage was repeated for years 9, 10 and 11.

The second topic concerned the recognition of position of a shadow thrown by an object. The correct response increased steadily from 6 years to 11 years. 10 per cent of boys gave correct responses at 6 years and 75 per cent at 10 and 11 years. The corresponding figures for girls were 22 per cent at 6 years and 61 percent at 11 years. At 6 years, 31 per cent and at 11 years 15 per cent of the children thought that the shadow was to be found between the source of light and the object. 52 per cent of the children at 6 years thought that the shadow was at right angles to the vertical plane through source of light and object, but this percentage decreased regularly to 17 per cent at 11 years. . . .

The next problem concerned positions on a seesaw if one of the two children was half the weight of the other. The lighter child was at one end of the seesaw and the children were asked to give the position of the other child as (*a*) at the other end of the plank, (*b*) at the fulcrum or (*c*) midway between the fulcrum and the other end.

The replies show a steady increase in percentages with age (25 per cent at 6 years to 71 per cent at 11 years). 67 per cent of the 6-year-old children gave the position of the lighter child as at the far end of the seesaw, but this percentage fell steadily and smoothly to 5 per cent at 11 years. Many children at ages 8 and 9 thought the position at the fulcrum gave the correct answer, although at 6 years only 9 per cent thought this was possible. Again, no significant difference appeared in boys' and girls' results. An interesting side light on the effect of experience on the results was provided by the comparison of the results from a small rural school and a large urban school. In the playground of the latter was a seesaw used by the children, yet 79 per cent of the children of the small rural school and 67 per cent of the larger school knew that the heavier boy must sit closer to the middle of the seesaw. A possible reason for this discrepancy was given by a member of staff of the rural school, namely, that the less formal atmosphere had enabled the children to improvise their own seesaws from equipment provided at the school.

Fig. 8: *Sand added to water.*



The companion question concerned a ladder carried by a man near one end and by another man near the middle. The children were asked to state who carried the heavier weight or did they carry the same weight?

Less than 10 per cent at all ages said that the two men shared the weight equally; 10 per cent to 15 per cent thought the man at the end of the ladder carried the heavier weight. Except for a difference at 6 years (girls 85 per cent, boys 73 per cent) both boys and girls gave 80 per cent to 86 per cent of correct responses at all ages.

The children were next reminded that ice can change to water and were asked, "which is the heavier—the ice or the water—or are they the same weight?" Less than 30 per cent at each age (except 11-year-old boys and girls) gave the correct answer. 80 per cent of boys and 65 per cent of girls believed ice to be heavier; these percentages fell steadily with increasing age until they reached a common value of 36 per cent. From the trends of the graphs it seems that complete realization of equality of weights does not occur until several years later. Most of the answers to the other parts of this schedule were based on direct observation, experience and even appropriate trial and error. Here was a question which required more accurate assessment than could be obtained by trial and error. The report of one school is relevant here. "Ice was stated to be heavier by almost half of 187 children tested (49 per cent). This was due to the fact that many children had some idea of the cause of burst pipes due to the water in them freezing to form ice. This ice, they had been told, took up more room than the water and caused the pipes to split open. There was an obvious confusion of volume and weight; as a boy of 10 years (M.A. 11.1) said afterwards, "If it is bigger, *it must* be heavier."

The next two questions concerned the movement of three cog wheels. A turns clockwise and the children were asked whether B or C or neither turned in the same direction as A.

The correct answer (C) showed a steady increase in response from 27 per cent at 6 years to 86 per cent at 11 years. After 8 years, more boys than girls gave correct replies (boys 90 per cent, girls 78 per cent at 11 years). The percentage of both boys and girls (64 per cent at 6 years) who gave B as their answer fell rapidly to 8 per cent at 11 years.

They were then asked which wheel turns most rapidly, A, B or C. Boys and girls gave the same pattern of response. 87 per cent at 6 years thought A, but this fell rapidly to 12 per cent at 11 years. There was a corresponding increase of correct replies (C) from 12 per cent (at 6 years) to 71 per cent (at 11 years). The graphs for speed and direction in these two related questions showed similar trends.

The last two questions in this section deal with two wheels connected, in the first case, by a direct belt and in the second case, by a transverse belt. In each the children were asked if the wheels rotated in the same or opposite directions. The graphs of percentages show a drop in correct responses between 6 and 7 years, and a corresponding increase in the wrong response. This may be due to the more practical and informal methods of the infants' classes contrasted with the change to the more formal type of work in the junior school. After the age of 7 years, the percentage of correct responses increases steadily to 90 per cent for both boys and girls. When the belt is twisted, the relative movement of the two wheels is not so obvious. 6-year-old girls do better than boys of the same age, but after 8 years boys do better than girls.

Section D

The first part of this section dealt with questions such as "Is the sun alive?" The same form of question was repeated in turn for tree, dog, flower, fire, candle, river, train, boat, aeroplane and bird. The children were instructed to give the answer as yes, no, or don't know.

That the dog and the bird were living gave the children no difficulty though 6-year-old boys (80 per cent) were a little more doubtful than 6-year-old girls (94 per cent). Succeeding ages gave a steady 98 per cent to 99 per cent response. The lower response of the boys was due to 13 per cent who were undecided. The graphs of the replies relating to boat, aeroplane, river and train were very similar. The majority (between 70 per cent and 90 per cent) after the age of 7 years said that these were not living things, and very few children at any of the ages failed to decide one way or the other. Sun and fire tended to have similar group responses with more variations between the ages and, in the case of fire, between the sexes. The children had great difficulty in deciding whether the sun was alive. At 6 years, 68 per cent thought it was and 25 per cent took the opposite view; at 9 years the two opinions were held equally strongly, but diverged again at 11 years (56 per cent said no and 35 per cent said yes). About 10 per cent of the children at each age could not give a direct answer. The candle tended to yield a consistent negative response, though 35 per cent of the children at 6 years thought it was 'living'; this shrunk to 11 per cent (boys and girls) at 11 years.

The replies to five of these questions tended to give reasonably steady results without much increase or decrease with increasing age. The replies to the questions whether a tree and a flower were living, as might be expected, gave similar patterns of response. More boys than girls were prepared to say that the tree was living until a common agreement of 92 per cent was reached at 11 years. Boys and girls agreed between themselves fairly closely and 93 per cent at 11 years thought that the flower was living; less than 10 per cent at any age were noncommittal. No criteria had been suggested either before or during the testing on what was to determine 'living' as opposed to inanimate things. There was slight evidence that movement was a factor by which the quality of 'living' was recognised by the majority of children in the age range but many were not prepared to accept this and looked for other criteria. The majority appeared to have come to a decision at least by 7 years, and for boat, aeroplane, river, train, candle, dog, bird, there was very little change in type of answer after that age; sun and fire gave more fluctuations than the other subjects over the age range between agreement and disagreement. Tree and flower gave increasing positive response with age with the realization that one of the main attributes of living things is the power to grow.

The last question summarised the preceding questions in this section. The children's answers showed the changes in thought content due to experience. 66 per cent at 6 years and 8 per cent at 11 years were prepared to state that all things that move are living. 28 per cent of the children at 6 years and 90 per cent at 11 years thought the opposite. If the responses of the few children in the 12-year age group are considered, it seems likely that at 12 years all (or nearly all) are prepared to deny that movement in itself is essential to life.

The confusion in the minds of the children roused by some of the earlier questions in this section was emphasised by the comments of those in one school as reported by their class teacher.

"Examination of the test material shows that many of the children who answered 'no' to the last question had earlier said that rivers were alive and, on discussion, they gave the reason that they moved. Others held that rivers were part of nature and were, therefore, alive, whereas a train was manufactured and was not living. Confusion here is helped by the fact that we speak of ships as 'she' and birds as 'it.' Matches are said to be *live* or *dead* and *live* coals fall from the fire in winter. We speak of *live* wires in an electric circuit, a meeting or discussion is *alive*, a town may have *life* or be *dead*.

Some of the children who answered 'don't know' to dog and flower explained that the dog or flower could be dead and, as it was not specified, they could not say a flower is still a flower after cutting. When it ceases to be a living flower is debatable. In one group of 11-year-olds the children discussed the difference between the life which makes us animate and the life which makes flowers grow."

The last part of this section deals with the child's written response to five questions only, two of which are discussed here. 1,386 answers to the question "What is the sky?" were received from children in the age range 6 years to 11 years. These were examined for content and at first grouped under eighty different headings. These varied from air, which was high up, thin and vacant, to the beginning of space or space which was around or formed a cover for the earth. Others mentioned various gases, presence of mist, water vapour, clouds, etc.; many stressed the colour or brightness, others the emptiness. One boy of 11 years (I.Q. 122) said it was an atmospherical illusion; a girl age 8 years 10 months (I.Q. 144) said "The sky is nothing really." The content of the answers ranged from single word descriptions of colour to adult responses involving attempts at scientific explanations. The various responses were then recorded under six headings: (*a*) True explanation; (*b*) false explanation; (*c*) true description; (*d*) false description; (*e*) fantasy; and (*f*) replies under the heading of 'don't know.' The responses fitted into these categories fairly readily although it is recognised that many different assessments of the answers are possible. In the 10- and 11-year groups more boys than girls attempted to give explanations and more girls than boys offered descriptive answers. If the answers of the small number of children age 6 years are omitted, all types of answer appear at each age; there is no evidence of definite stages. There may be stages in the growth and application of the thought processes which are latent in the summary of total replies in Table 2, but in general, there is a slow fall in descriptive answers and a gain in the number of attempts at explanations with age: 43.6 per cent gave true answers either in form of description or explanation. 43.0 per cent gave false answers including fantasy. 13.3 per cent gave no reply or answered 'don't know.'

The second question asked "What is the night?" This involved a great variety of responses, which were tabulated under the headings used for the first question. [See Table 3.] On the whole there were fewer answers that made use of fantasy and more answers that made use of true explanations or descriptions. It appears that the number of true descriptions decreases with age and this type of answer is replaced by a steady increase in attempts to explain in terms of movement of earth, etc.; failure to answer decreased as age increased and answers that involved fantasy were less than 6 per cent at any age.

Table 2: "*What is the sky?*"

	6	7	8	9	10	11	Nos.	%
True explanation	0	0.7	3.2	7.2	19.8	22.4	165	11.9
False explanation	6.7	3.2	2.3	5.6	9.6	5.8	83	6.0
True description	56.7	31.2	33.3	29.1	31.1	29.0	440	31.7
False description	30.0	34.1	27.0	28.1	24.0	24.5	371	26.8
Fantasy	3.3	9.4	15.3	14.1	6.9	8.3	142	10.2
Nil (don't know)	3.3	22.4	18.9	15.9	8.6	10.0	185	13.3
						Total	1386	

Table 3: "*What is the night?*"

	6	7	8	9	10	11	Nos.	%
True explanation	0	1.4	4.5	15.9	23.2	40.2	254	18.3
False explanation	0	1.4	1.4	5.6	7.4	14.9	89	6.4
True description	75.0	54.3	59.5	49.1	43.0	29.9	655	47.3
False description	18.3	22.5	18.9	16.9	14.3	7.1	213	15.4
Fantasy	1.7	0.7	1.4	1.6	2.7	0.8	23	1.7
Nil (don't know)	5.0	19.6	14.4	10.9	9.4	7.1	152	11.0
						Total	1386	

True explanations or descriptions were found at all ages (except in the small sample of 6-year-olds). Here are some of the replies:

Girl C. A. 6.0: "When you sleep, when the stars shine, when the moon shines."

Boy C. A. 7.2: "Black."

Girl C. A. 8.0: "Night is after or before day. It is usually dark."

Girl C. A. 11.9: "Night is the time when the earth has spun round so that where you are does not face the sun, when it has spun on its axis so far darkness descends."

The first three answers were classified as descriptive and true in meaning as far as they went. The last was considered an attempt at a true explanation for a child at this age.

Obviously these two questions would have a different impact on the children and would affect them differently. Sky and night are both familiar in the sense that they are experienced every day, but night is nearer and has a more vital impact on their habits. This is shown by increase in percentages for attempts at true explanations and descriptions and the corresponding fall in answers that border on fantasy and the complete absence of any religious explanation.

The greater percentages of correct replies to both questions obtained by the 6-year-olds compared with results from other age groups is due entirely to the fact that they gave responses wholly in terms of simple descriptions and had not reached a stage of answering in terms of why and because. Not until 10 years of age did a quarter of the sample of children offer explanations.

CONCLUSIONS

When the results are reviewed as a whole, two types stand out clearly. Altogether, twenty-four answers from boys and girls show steady increase with age; eighteen of these are given in Table 4, with relevant percentages of correct answers for each of the six years. They do support the thesis advanced by Deutsche (1937, p. 33) that they justify the method of presentation of material and the method of scoring. If either or both were inadequate, such increase with age would hardly be obtained. Inspection of these twenty-four topics shows fairly clearly what they all have in common; these questions could be correctly answered by experience without formal teaching and by observation in and outside the classroom. Where questions in this category allowed the children to record replies under the heading 'don't know,' the percentage of children who did so was small (usually less than 10 per cent).

Of the remaining questions, thirteen showed definite confusion and very little increase of correct response with age. Estimation of length, weight direction and volume were difficult even for adults. The association of volume and weight was not understood. Topics like these cannot be understood solely by experience; a more formal scientific approach to units of length, weight and volume (most of the younger children had very little experience with these units) is needed to supplement experience. Without formal demonstration of some kind, it cannot be assumed that children know the constancy of volume or weight or the change of volume when ice turns to water.

It is clear that in the acquisition of some skills, maturation may be sufficient but in others formal teaching helps, but it is also true that the evidence in this research suggests that the degree of difficulty determines the kind of response and, in fact, these questions proved difficult when formal explanations were not present to supplement experience.

The section involving the distinction between animate and inanimate things showed confusion in certain specific items (e.g., sun and fire), but movement was recognised as not the sole criterion of life and the scores 'for and against' at all ages showed a remarkably smooth distribution; so far as the answers of the group were concerned, there was no evidence of Piaget's stages of development but only a gradual development of the reasoning processes by more systematic organisation of concepts.

Where children had the opportunity to give free answers to the questions on sky and night, types of responses were spread through all the age groups and no one age group had a monopoly of answers of a given type. Scientific explanations, verbal descriptions, religious references, naive conjections could be found at all ages for boys and girls. Experience in and out of school and a vocabulary increasing with age seemed to be the main factors that determined the types of answers given by boys and girls to these two questions. This agrees with the study by Oakes (1947) who found no evidence "to corroborate Piaget's interpretation that there is a definite stage in the child's thinking which is characteristic of a given age." There is little doubt that the types of response (causal explanation, descriptive, enigmatic, religious) is determined by the question, and this is just as likely to be true of adults as of children; to the question about the sky 10.3 per cent of the total children gave answers not classified as explanatory or descriptive but to the question 'what is the night?'

Table 4: *Table showing percentage of correct replies and chronological age*

	Age in Years					
	6	7	8	9	10	11
1—A.1. Estimate 15 secs.	8	34	43	49	55	67
2—B.1. Stone added to water	54	70	74	86	89	88
3—B.3. Level of water in second flask	44	61	67	73	86	91
4—B.7. Plasticine : wt. and shape	43	48	63	64	78	79
5—C.1. Steam = water	41	37	51	74	78	88
6—C.1. Hail = water	30	49	59	74	76	77
7—C.1. Clouds = water	41	40	50	56	64	71
8—C.2. Shadow of vertical post	18	36	48	53	66	67
9—C.3. See-saw	25	38	47	55	68	71
10—C.6. Direction of cog wheels	27	39	56	64	76	86
11—C.7. Speed of cog wheels	12	21	42	52	67	71
12—D.1. Flower-living ?	58	49	67	78	88	93
13—D.3. Are all things that move living ?	29	42	58	74	84	90
14—E.2. Cork and needle	28	31	50	73	71	81
15—E.4(d). Shadow of disc—as edge	21	27	37	39	42	56
16—E.5(e). Section through centre of sphere	42	38	58	68	64	64
17—E.5(f). Section of cone perpendicular to axis	6	12	34	41	43	61
18—E.5(g). Section of cone perpendicular to base	34	30	47	66	60	63

only 1.7 per cent answered in this way. In both cases, with increase of age, reasoning gradually replaced descriptive words, percepts gave way to concepts. There is evidence that children behave like adults, that is they lack in their early years the correct words and use a limited range of words to verbalise in the same way as adults would meet the same difficulties in describing or attempting to understand some of the modern scientific discoveries. As D. H. Russell (1956) reported "In some experiments, children show clear understanding of a concept but inability to verbalise it."

Perhaps there is a need about 8 or 9 years of age for the harnessing of this experience so that knowledge can be obtained by other, though not necessarily more formal, means. These experiments have shown that children of Primary school age have accumulated a knowledge of scientific facts and that this knowledge increases with age in some cases very steadily. Some of the answers had to be obtained by reasoning on the bases of past experience and these also showed steady increase with age up to 8 or 9 years. Beyond that, *a priori* reasoning was not sufficient and adult guidance and explanation seemed essential.

REFERENCES

DEUTSCHE, J. M. *The development of children's concepts of causal relations*. Minneapolis: University of Minnesota Press, 1937.

HAZLITT, V. Children's thinking. *Brit. J. Psychol.*, 1930, *87*, 447–531.

ISAACS, SUSAN. *Intellectual growth in young children*. London: Routledge, 1930.

JOHNSTON, E. C., & JOSEY, C. C. *J. abnorm. soc. Psychol.*, 1931, *26*, 338–339.

OAKES, M. E. *Children's explanations of natural phenomena*. New York: Teachers' College, Columbia University, 1947. Contr. Educ., No. 926.

PIAGET, J. *The child's conception of the world*. London: Routledge & Kegan Paul, 1929.

PIAGET, J. *The child's conception of physical causality*. London: Routledge & Kegan Paul, 1930.

RUSSELL, D. H. *Children's thinking*. Boston: Ginn & Co., 1956.

40

The Development of Conscience: A Comparison of American Children of Different Mental and Socioeconomic Levels

Leonore Boehm

Boehm follows up the work Piaget has done on the development of conscience. Her investigation is descriptive in that it presents the findings in relation to already existent cultural and intellectual variables. A similar approach is found in the article by Johnson (59). For contrast, in terms of an experiment designed to modify development, see Ojemann and Snider (33).

This is [a report] on an investigation of moral judgment and conscience development, and subcultural differences due to religion, socioeconomic class, and mental level. The writer agrees with Piaget (1932) that the cognitive ability to distinguish between right and wrong is one of the ingredients in conscience development, and she was interested in investigating at what age different groups of American children learn to distinguish between intention and result of an action. Another closely related question arose from Piaget's belief, which the author questioned, that the child cannot attain "morality" until he becomes independent of adults and achieves peer reciprocity.

The subcultural comparisons were made by studying boys and girls of two intelligence levels in upper middle class and working class attending public and parochial schools. The parochial schools were Jewish and Catholic.

After having interviewed a number of children, the author formulated the hypotheses that, in answer to questions concerning intention versus result: (a) academically gifted children might respond with earlier maturity than children of average intelligence and (b) upper middle-class children might respond with earlier maturity than children of working-class parents; but that, in answer to questions regarding independence from adult versus peer reciprocity, (a) gifted children might be less mature than children of average intelligence and (b) upper middle-class children might be less mature than working-class children. A further hypothesis was that

---

From *Child Development*, 1962, *33*, 575–590. Reprinted with permission of the Society for Research in Child Development.

Table 1: *Population: children by school and group*

Group	Total	SCHOOL			
		Public	Catholic Parochial		Jewish Parochial
			Natick	*Brooklyn*	
Total	237	112	44	66	15
Gifted					
Upper class	80	22	20	23	15
Working class	43	19	8	16	..
Average					
Upper class	56	34	8	14	..
Working class	58	37	8	13	..

there would be individual and subcultural differences in the age of mature response, depending upon the area of conscience development under investigation and specific aspects of that area.

PROCEDURE

Table 1 shows the distribution of children interviewed in the study according to above-average intelligence (IQ over 110) and average intelligence (IQ 90 to 110), socioeconomic class, and the school attended.[1] These children ranged from 6 to 9 years of age. It was decided to set an upper limit of 9 years, because analysis of data in the previous study (Boehm & Nass, 1962), which had included children from 6 to 11 years, indicated little change in response after the age of 9.

The children comprised all students available in their grades who fell into the desired categories, except in the Jewish parochial school where the principal selected the small number of children we were allowed to interview.

Warner's norms were used as a basis for assigning social class position of the subjects studied.

Using Piaget's "clinical method," each child was seen individually and told four stories. Two stories were by Piaget, the Cup story with slight cultural adaptation and the Lost story. The other two stories (Boehm, 1957), Fight and Scout, had been constructed and tested in collaboration with Alina Szeminska. Interviews were recorded on tape. The Cup and Lost stories are concerned with the evaluation of an act as to intent or result, while Fight and Scout are concerned with peer reciprocity versus dependence on adult authority.

The stories were used as semiprojective tests. It was hoped that the children tested would identify with the children in the story and would respond on an affective, as well as on a purely intellectual level.

The stories are presented below with questions illustrating the type of question used in the interviews. All but the Lost story were interrupted by questions before the stories were completed. In the clinical method the investigator formulates each question on the basis of the response to the preceding question and a uniform questionnaire cannot be used. In the course of probing, the investigator asked a large number of questions, making sure to pick up the child's exact terminology in order to preserve the child's emotional reaction.

CUP STORY (after Piaget):

A little boy named John (or a girl named Mary, depending on the sex of the subject) is called to dinner by his mother. Between the room where John is and the kitchen (or dining room) there is a swinging door. Behind the door is a chair with fifteen cups on a tray. John has no way of knowing that the chair with the cups is there. When he opens the door, the door swings into the chair, the chair and the tray tumble and all the cups break.

How do you think his mother felt about this? Was she angry at John?

---

[1] The investigator realizes that the samples when grouped by age, intelligence and class, as well as by type of school attended, are small; however, age development can be observed when all children of a given age are grouped together.

A boy named Henry (or a girl named Margaret) is told by his mother not to take any cookies from the jar, since the mother needed the cookies for guests. The mother leaves the house and Henry wants to snitch one cookie. Since the mother has placed the jar out of his reach, Henry has to climb up on a chair in order to reach the jar and he knocks over a cup on a lower shelf while doing so.

How will Henry's mother feel when she comes home?

Which mother is more angry, John's or Henry's?

What do you think about these two boys?

Are they both naughty? Is one naughty?

If you were the father (or mother) would you punish both?

Would you punish one more than the other? Which one?

LOST STORY (after Piaget):

A boy named Joe (or a girl named Alice) has just moved into the neighborhood and does not know the names of the streets very well. One day while he is playing, a man stops him and asks him where ——— Street (street near child's school) is. Joe is not sure of the direction, but answers: "I think it's over there." But it was not there. The man walks and walks and walks and cannot find the house he is looking for.

A boy named Mike (or a girl named Louise) knows the neighborhood very well. One day a man asks him where ——— Street is. Mike knows, but he wants to play a trick on the man and says, "It's there," and shows him the wrong way, but this man is smart and finds his way.

What do you think of the two boys?

Why?

Were both these boys equally naughty?

FIGHT STORY (Boehm and Szeminska):

Two boys, Louis and George, very good friends, want to find out whether one of them is stronger than the other. So they decide to have a match, a fun fight, to see who is stronger. The next morning they meet early in the school yard for their fight. While they are fighting, by accident Louis hits George on the nose which begins to bleed badly.

How do you think George felt about Louis?

How did Louis feel?

Because Louis felt sorry (or whichever term the child used) he wanted to do something about it, and wondered what to do. He could ask his teacher. Louis thought the teacher might tell him to write "I should not fight," 100 times. Or he could ask some of his friends. One of them might say, "Why don't you bring George your favorite toy to play with?" Another one might say, "Buy him a gift." A third might say, "Just go to George and tell him you're sorry."

What do you think Louis did?

Why?

Louis went to George to tell him he was sorry (using the child's term) and he also brought him his toy. But George said, "You don't need to give me anything. You didn't mean to hurt me. It was an accident. I could have hurt you just as easily."

Do you think Louis left the toy or took it home?

Why?

When do you think George forgave him, when Louis said he was sorry, or when he gave him the toy?

Would the toy make George be a better friend again?

What do you think made Louis feel better, writing 100 times "I should not fight," giving his toy to George, or hearing George say everything was all right?

SCOUT STORY (Boehm and Szeminska):

A group of children X years old (the subject's age) want to give a surprise birthday party for their scout leader. One boy, Charles, is to decorate the room. He has never done this before and wonders whom he should ask for advice. He could ask his home room teacher. This teacher does not teach arts and crafts because he is not good in it, although he is very good in English, social studies, mathematics, etc. Or Charles could ask a boy in his classroom who takes lessons in arts and crafts at the museum once a week and is excellent. He is so talented that he has won a prize at the museum.

Whom do you think he asked?

If he asked both and they gave ~~him~~ different ideas, whose advice would he follow?

If he thought both ideas were equally good, whose would he use?

A friend told him he liked the classmate's plan much better than the teacher's. Would Charles still rather use the teacher's?

If he follows the friend's advice and uses the boy's plan, will the teacher be hurt or angry?[2]

TREATMENT OF DATA

Each child's protocol was scored independently, from transcriptions of the tape, by three judges, one of whom was the investigator. The other two judges were graduate education students, both teachers. Responses to each story by each child were classified according to Piaget's three levels of morality (1932):

> 1. "Morality of constraint." This is defined to mean automatic obedience to rules without reasoning or judgment, with adults viewed as omnipotent. Punishment is seen as a necessary retribution of justice and given in proportion to size or result of the misdeed, independent of motive. Intentions are not taken into consideration.
> 2. An intermediary stage in which the child internalizes rules without evaluating them, or alternates in his responses to the situation.
> 3. "Morality of cooperation." This is defined to mean that the child evaluates intentions rather than deeds or outcomes alone. Moral behavior is engaged in for its own sake, not out of fear of punishment. Cooperation and mutual respect among peers have taken the place of former unilateral respect for adults.

Because of the relatively small numbers of responses, it was not expedient to examine responses "1" and "2" separately in the statistical analyses.

RESULTS

Our findings were analyzed by two different statistical methods: (a) tests of interactions (Table 2) and (b) significance tests on various proportions. In the tests of interactions the responses (omitting the small Jewish group) were classified according to three independent variables (mental level, socioeconomic level, and stories) and their combinations.

Table 2: *Tests of interactions*

Classifications	*df*	x^2
Mental level	1	15.03*
Socioeconomic class	1	2.11
Mental level by socioeconomic class	1	11.49*
Stories	3	30.20*
Stories by mental level	3	1.40
Stories by socioeconomic class	3	8.11*
Stories by mental level by socioeconomic class	3	.56

* $p \leq .05$.

Both types of analysis show that mental level is significant and is modified by class membership. When in the tests of interactions the results of all four stories are combined, socioeconomic class differences are masked, because the pairs of stories operate in different directions concerning class level. When considered individually, however, the stories are sensitive to socioeconomic class membership. Although the four stories vary in difficulty, their relative difficulty does not vary for the different IQ groups.

In each pair of stories one story received fewer "3" responses than the other. Cup and Fight presented much greater difficulties to the respondents, whose scores on both stories were lower than expected on a chance basis. Cup scored lowest, while Scout and Lost exceeded the expected number of "3" responses, Lost more than Scout.

The percentage of "3" responses to the Cup and Lost stories (concerned with intention and result of action) and to the Fight and Scout stories (concerned with adult authority and peer reciprocity) are not significantly different: 62 per cent in the case of the former, and 59 per cent in the case of the latter.

Responses to the stories were compared according to age groups. Gifted children gave "3" responses at an earlier age and at each age level received more "3" scores than children of average intelligence, except for the Scout story, when the socioeconomic classes were combined.

These differences between gifted children and those of average intelligence were accounted for mainly by the upper middle-class group. While the gifted upper middle-class children gave significantly more "3" responses (except for the

---

[2] These stories are in a slightly altered form than when first used.

Scout story) than subjects of average intelligence in the upper middle class, gifted working-class children (in all but the Scout story) showed only a slightly higher (and nonsignificant) proportion of "3" responses than their counterparts of average intelligence.

Responses to the separate stories were compared by socioeconomic class. In the Cup and the Lost stories, in which children evaluated a deed as to either intention or result, the upper middle class, of both levels of intelligence, gave more "3" answers at an earlier age. In the Cup story there was a larger difference favoring the upper middle class among gifted children than among those of average intelligence. Among the latter at ages 7 and 8, there were small but nonsignificant differences in favor of the working-class group.

The Lost story also showed higher scores for upper middle class with all ages combined, although at ages 6 and 9 gifted working-class children scored more "3's" than gifted upper middle-class children. At ages 7 and 8, working-class children of average intelligence scored more "3's" than upper middle-class children of average intelligence. These differences in favor of working-class children were less marked than the ones in favor of upper middle-class children.

For the Fight and the Scout stories which are concerned with peer reciprocity versus dependence on authority, the working-class child at both intelligence levels scored higher than the upper middle-class child. In the Fight story, with a few exceptions, the working-class children tended to give more "3" responses than middle-class children of comparable age and intelligence. In the Scout story the scores are slightly, though insignificantly, in favor of working-class children at both intelligence levels when ages are combined.

The age at which a majority of the children tested gave level "3" responses varied with the specific situation and the child's socioeconomic class and level of intelligence. The response to the Cup story by gifted children showed that a majority in the upper middle class gave "3" level responses at age 6, while the majority of the working class did so at age 9. Among children of average intelligence, there was no class difference, the majority scoring "3" only at 9 years.

Level "3" responses were given to the Lost story at age 6 by a majority of the gifted and average upper middle-class children and the gifted working-class children. The majority of working-class children of average intelligence were 7 years old before they gave a "3" level response.

In the Fight story, the gifted group in the upper middle class scored "3" at age 7, and the gifted in the working class at age 6 (although there is an exception at age 7). In the group of average intelligence, the upper middle class scored "3" at age 9 and the working class at age 8.

In the Scout story a majority of "3" answers were given at age 6 by gifted upper middle class, at age 7 by gifted working class, and at age 8 by upper middle class of average intelligence. With the responses available, it was not possible to determine a "3" level response for the working-class child of average intelligence since, although the majority scored "3" at ages 6 and 8, at age 7 only 47 per cent scored "3."

Gifted children scored "3" responses at least a year earlier than those of average intelligence, except for the Scout story.

The majority of gifted children in both socioeconomic groups scored "3" at age 9 for the Cup story, at age 7 for the Fight and Scout stories, and at age 6 for the Lost story.

The majority of children of average intelligence in both socioeconomic groups scored "3" at age 9 in the Fight and Cup stories, at age 8 in the Scout story, and at age 7 in the Lost story.

The findings regarding class and intelligence show that, in stories concerned with motivation versus outcome, gifted upper middle-class children had the greatest relative number of "3" scores. They were followed by upper middle-class children of average intelligence. Gifted working-class children received somewhat fewer "3" scores than upper middle-class children of average intelligence. Working-class children of average intelligence had the lowest number of "3" scores.

In stories concerned with the child's independence from adult authority, the picture is somewhat reversed. Gifted working-class children had significantly more "3" scores, followed by gifted middle-class children. Working-class children of average intelligence had more "3" scores than upper middle-class children of average intelligence, the difference approaching significance.

DISCUSSION

These results generally confirm the hypothesis of a difference between gifted children and those of average intelligence with regard to maturity of moral judgment (as defined by Piaget).

A review of the literature on this subject discloses that other investigations did not find this distinction. The findings of these studies, which are also built on Piaget's method and concepts, are at variance with the present findings because of such factors as the smaller number of subjects at each age, differences in range of age groups, differences in range of IQs in the gifted and average categories, and in the specific aspect within the area of moral development being investigated. . . .

Durkin (1959a, b, and c) used Piaget's clinical method in interviewing boys and girls in grades 2, 5, 8, and 11 on the development of concepts of justice. In her first study (1959a) of fifth grade children, Durkin felt there was a possible difference in level of responses due to intelligence level; children of higher intelligence were less inclined to seek help from an authority figure than those of lesser intelligence. Older and younger subjects, regardless of intelligence level, however, preferred to report an incident to a person in authority rather than settling it among themselves.

In a later paper Durkin (1959c) concluded that this difference among fifth grade subjects was due to chance and that there was no difference in response between intelligence levels. Several factors may account for the fact that her findings differ from those of the present study. The subjects from a small midwestern, mostly farming, community, though considered by Durkin to be middle class, were a mixed group economically and from a lower middle-class group than those in the present study. The IQ scores of her subjects also differed, ranging from 92 to 114 in the second grade, from 69 to 122 in the fifth grade, and from 77 to 148 in the eighth. Subjects of above-average intelligence in the present study had IQs higher than those in Durkin's above-average group.

A few of the gifted children in the present study, like those in Durkin's, chose adult help over peer help, but they obviously were trying to please the investigator with their replies. They "changed" their opinions at the slightest suggestion from the interviewer. Bright children are more likely to try to please teachers, since they have more foresight and are aware of possibilities for ingratiating themselves with persons in authority. Often they identified with the interviewer who represented a teacher to them and gave the answer they thought was expected.

MacRae (1950, 1954) investigated Piaget's concept of morality, using Piaget's stories in interviews with younger children and in questionnaires filled out by older children. His subjects from grades 5 to 8 consisted of a small group of boys of higher IQs (over 105) who were compared with those of average intelligence, each group being composed of higher and lower socioeconomic classes.

MacRae found more adult dependency among the gifted children in the fifth grade. MacRae's fifth-graders of above-average intelligence believed more in immanent justice or punishment and thus judged more absolutely and rigidly than those of average intelligence in the same grade. His gifted and upper-socioeconomic-class boys allowed no exception in breaking of norms, even in circumstances where a violation of a norm would appear more moral than keeping it.

In this respect, his older subjects responded alike without regard to intelligence levels, but at all ages children of above-average intelligence and those of upper socioeconomic class responded more maturely to questions concerning intention and result and took into account foreseeable consequences of actions. They responded more maturely with regard to strict punishment and in showing they knew when they were capable or not to judge.

MacRae decided that children of above-average intelligence and upper socioeconomic class respond more maturely to "cognitive" aspects of moral development. Cognitive responses, according to MacRae, are based on knowledge of adult expectations and are due to cultural indoctrination and learning of norms. He believes that the bright child and the child of upper socioeconomic class internalizes parental rules more strongly and remains dependent upon parents longer than those of average intelligence and the lower class. Because of this, they cannot be flexible about rules and will not accept any change in rules. The present findings do not bear out MacRae's findings of longer or stronger dependence on adults by gifted children than by children of average intelligence. Gifted children at almost all ages—there are exceptions in the Scout story—responded with less adult dependency than those of average intelligence.

Differences in responses between gifted upper middle-class and gifted working-class subjects are less surprising than would first appear. Piaget pointed out that children's actions are "moral" far earlier than their theoretical verbal judgments. On this point MacRae's (1950) findings and conceptions of moral evaluations and adult independence in the working-class group may well be correct. These children, regardless of intelligence level, are certainly less accustomed to formalized thinking. The less verbally inclined, though gifted, working-class child could be as "moral" in his actions as gifted upper middle-class children, but might appear less mature because of an inadequate way of expressing himself in judging actions. However, in his responses it was not only the quality and length of expression which differed, but the level of moral judgment.

In calling for a distinction between intention and consequence, the Cup and Lost stories could have involved a higher thought process than the other stories, a more complex "operation" (Piaget) and a more conceptually abstract thought. Both the Cup and Lost stories, in contrast to the other two, contain a pair of plots similar in some respects and different in others. However, more younger children from both levels of intelligence distinguished between intention and result in the Lost than in the Cup story, although both stories are concerned with the same problem.

In the Cup story, working-class children often judged "maturely" (according to Piaget's definition) after hearing the first plot but changed to an "immature" response when hearing the second plot and being asked to compare them. While at first they were able to judge the action without regard to result, in the second part of the story they evaluated outcome, impressed by the comparison of 15 versus one broken cup. Perhaps they would have remained more objective if the number of broken cups had varied less. As it was, emotions might have interfered with logical thought. Personal experience with scoldings and punishment for dishes broken accidentally might be remembered. A belief in "immanent justice" returned and with it the expectance of dire results as a consequence of behavior causing harm.

Contrary to Piaget, our data do not show that "maturity of moral judgment" increases as the child becomes independent of adults and achieves peer reciprocity. In the present study there seemed to be no connection between "mature" responses concerning right and wrong and responses indicating adult independence. Rather, working-class children who gave early "3" level responses to the stories concerning adult independence scored more lower level responses at the same age when distinguishing right from wrong in considering intention and result of actions. Upper middle-class subjects, who were more discerning at an earlier age than working-class children with regard to intention and consequence, scored more lower level responses at the same age regarding issues of adult independence and peer reciprocity.

SUMMARY AND CONCLUSIONS

In a study of some aspects of conscience development in public schools, two Catholic parochial schools, and one Jewish parochial school, academically gifted children were compared with children of average intelligence, and upper middle-class children with children of the working class. The following findings emerged:

1. Academically gifted children mature earlier in their moral judgments concerning distinctions between intention and outcome of an action than children of average intelligence.
2. Upper middle-class children develop earlier in this aspect than working-class children.
3. There is a greater difference between responses of academically gifted children and children of average intelligence of the upper middle-class than between gifted children and those of average intelligence in the working class.
4. Working-class children at both intelligence levels show earlier peer reciprocity and adult independence than upper middle-class children.

REFERENCES

BOEHM, LEONORE. The development of independence: a comparative study. *Child Develpm.*, 1957, *28*, 85–92.

BOEHM, LEONORE, & NASS, M. Social class differences in conscience development. *Child Develpm.*, 1962, *33*, 565–574.

DURKIN, D. Children's concepts of justice. *Child Develpm.*, 1959, *30*, 59–67. (a)

DURKIN, D. Children's acceptance of reciprocity as a justice-principle. *Child Develpm.*, 1959, *30*, 289–296. (b)

DURKIN, D. Children's concept of justice. *J. educ. Res.*, 1959, *52*, 252–257. (c)

MACRAE, D. The development of moral judgment in children. Unpublished doctoral dissertation, Harvard University, 1950.

MACRAE, D. A test of Piaget's theories of moral development. *J. abnorm. soc. Psychol.*, 1954, *49*, 14–18.

PIAGET, J. *Le jugement moral chez l'enfant*. Alcan, 1932.

41

Studies in Concept Attainment: III. Effect of Instructions at Two Levels of Intelligence

Sonia F. Osler and Sandra Raynes Weiss

Osler and Weiss examine the interaction between age and intelligence and the type of instruction received in learning a concept. According to Bruner (8), one of the factors influencing concept attainment is whether or not the subject knows that there is a concept; the research by Osler and Weiss clearly illustrates this factor. A study that centers on the importance of the type of instruction has implications for teachers and parents alike. See also Ojemann and Pritchett (24) and Saugstad and Raaheim (61).

In a recent study on the effect of intelligence on concept attainment using instructions which were nonspecific as to the nature of the task, Osler and Fivel (1961) reported a significant difference in errors between groups of average and superior IQ. An important question that has yet to be answered concerns the pertinent attribute of high intelligence which accounts for the superiority of performance. One possibility is that superior *S*s supplement *E*'s instructions with their own, directing them to search for consistencies in the reinforced stimuli; whereas the less intelligent *S*s work along without self-instructions until the reinforcement contingencies of the experiment strengthen the response to the concept exemplars. Perhaps a simpler statement of this hypothesis might be that under the conditions of nonspecific instructions the task presented to *S* has two components, namely, problem *finding* as well as problem *solving;* and that the superior performance associated with high intelligence may be the result of greater proficiency at either or both of these components. On the other hand, when instructions are specific, defining the problem which *S* is expected to solve, the problem finding component of the task is eliminated or reduced to a minimum.

It was the aim of the present investigation to test these ideas experimentally. Two experiments were performed. Experiment I was a replication of the Osler and Fivel (1961) experiment. This was undertaken for the double purpose of confirming the earlier data and controlling for the possibility that automated equipment presently being used for the conduct of the experiment may have affected the results obtained. In Experiment II the general instructions were replaced by more specific ones, defining the problem for *S* and directing him to search for hypotheses.

METHOD

As Exp. I and II differed only in the instructions used, the method for both will be described together, and wherever necessary information pertinent to each experiment will be provided separately.

Subjects

The *S*s consisted of 180 school children attending elementary and junior high schools of Baltimore City and County, equally divided among the following age groups: 6, 10, and 14 yr. Each *S* was within 6 mo. of the age specified. Half of the *S*s within each age level had a WISC IQ between 90 and 109 (Average Intelligence group), and the other half had a WISC IQ of 110 and above (High Intelligence group). The mean IQs of the two groups in Exp. I were 100.8 and 120.9, and in Exp. II they were 101.2 and 121.7, respectively.

Apparatus

The apparatus has been described elsewhere (Osler & Powell, 1960). Essentially, it consisted of a film strip projector and a screen to receive the projected images of the stimuli which appeared in pairs. A telegraph key placed under each stimulus released a marble following a correct response, which consisted of pressing the key under the concept exemplar. No reinforcement followed if the key under the negative stimulus was pressed. The *S*s were not paced, but 4 sec. after a response was made the next stimulus was presented.

The concept studied was *bird.* This was one of the three concepts on which data were reported by Osler and Fivel (1961). The stimuli used were the same as in the previous investigation, except that they now appeared on a

---

From *Journal of Experimental Psychology,* 1962, *63,* 528-533

screen projected from a film strip. The complete set of stimuli consisted of 150 pairs of pictures, one member of each pair representing a positive instance of the concept and the other a negative instance. The pictures of each pair were matched approximately in area, color, and brightness. Reasons for using naturalistic stimuli, which carry with them the hazard of unspecified dimensions, have been discussed elsewhere (Osler & Fivel, 1961). An additional reason prompting their use in this investigation was the previous finding that performance on this concept was a function of intelligence, thus demonstrating the suitability of this problem for the present study. The stimuli had been pretested on a group of children of the same age and intelligence levels as the present *S*s, and were found to be familiar to the type of subject studied.

Procedure

Each *S* was tested individually in two separate sessions. At the first session *S* was told that he would play a game in which he could win any one of a group of toys displayed before him. He was asked to select the most preferred of the group, and this was then set slightly apart from the others in a position which allowed *S* to view it during the "game." The *S*s in Exp. 1 received the following instructions:

> This is the game we are going to play. Here are two handles (*E* points to them). If you push the correct handle you will get a marble. Sometimes this will be the correct handle and sometimes this will be the correct handle. If you push the correct handle you will get a marble like this (demonstrated). If you push the wrong handle, like this, no marble will come (demonstrated). If you get a marble you can put it here (here *E* points to a grooved wooden tray with which *S* was provided). Now you try this one and let's see what happens. If you look at these pictures before pushing the handle, you will be able to get many marbles. When you fill the tray with marbles, you can have the (name of the toy).

The stimuli used during the demonstrations were unrelated to the concept.

In Exp. II the *S*s were also informed about a game in which they could win a toy, and then they received the following instructions:

> This is the game we're going to play. Listen carefully and I will tell you how to play the game. Here are two pictures. You pick one of them and push the handle under it, and if you are right, you will get a marble, like this (demonstrated). If you are wrong you will not get a marble, like this (demonstrated). Now you try and push this one. The game is to see how soon you can get a marble every time you choose. If you look at the pictures carefully you will see that there is something in the pictures like an idea that will tell you which one to choose to get a marble every time. Now we will start the real game. What are you going to do first? (If *S* does not say: look at the pictures, *E* supplies the answer). If you look carefully and see which pictures give you the marble you will be able to win marbles every time. Are you ready? Remember, look at the pictures first. (The preceding sentence was repeated together with the following sentence whenever necessary.) Remember, there is an idea that will tell you which picture will give you the marble.

The procedure was continued until *S* made 10 consecutively correct responses, which was the criterion of success. At this point *E* commented on *S*'s good work and asked him how he knew where the marble was each time. If *S* could not state the rule correctly the procedure was continued with questions following each block of 10 consecutively correct responses until *S* gave a correct answer or completed 150 trials. Any *S* who did not achieve criterion performance at that point was considered to have failed. A second session, held within 2 weeks of the first, was devoted to the administration of the WISC.

Experimental Design. In both experiments a 3 × 2 factorial design was used with three ages and two intelligence levels constituting the respective variables. Thus there were six experimental groups with 15 *S*s in each group, a total of 90 *S*s in each experiment.

RESULTS

The results were analyzed in terms of three measures of performance: (*a*) the mean number of errors to criterion, (*b*) the number of *S*s who attained the criterion, and (*c*) the number of *S*s who verbalized the concept at the point of success.

Experiment I.—The mean number of errors per group and the results of an analysis of variance of errors, presented in Table 1, indicate that age and intelligence were associated with reliable differences in performance, thus reproducing

Table 1: *Mean number of errors to a criterion of 10 consecutive correct responses with general instructions*

Intelligence	GROUPS						Means by IQ
	6-Yr. *Ss*		10-Yr. *Ss*		14-Yr. *Ss*		
	Mean	*SD*	Mean	*SD*	Mean	*SD*	
Average IQ	71.2	21.0	38.9	34.5	14.3	22.2	41.5
High IQ	57.1	27.8	19.7	17.0	6.5	11.4	27.8
Means by Age	64.2		29.3		10.4		

Note.—$F_{Age} = 37.66$, $df = 2$, $P < .001$; $F_{Intel.} = 7.14$, $df = 1$, $P < .01$; $F_{Age \times Intel.} = < 1$.

Table 2: *Number of Ss who reached criterion and number who verbalized the concept under general instructions*

Intelligence	GROUPS						All Ages	
	6-Yr. *Ss*		10-Yr. *Ss*		14-Yr. *Ss*			
	Crit.	Verb.	Crit.	Verb.	Crit.	Verb.	Crit.	Verb.
Average IQ	4	1	8	2	13	13	25	16
High IQ	7	4	14	11	15	14	36	29
Total	11	5	22	13	28	27	61	45

Note.—Criterion data: Age $\chi^2 = 22.690$, $df = 2$, $P < .001$; Intelligence $\chi^2 = 6.156$, $df = 1$, $P < .02$. Verbalization data: Age $\chi^2 = 33.068$, $df = 2$, $P < .001$; Intelligence $\chi^2 = 7.512$, $df = 1$, $P < .01$.

the findings reported by Osler and Fivel (1961). Similar results were obtained from an analysis of the data on frequency of *S*s who attained criterion performance and those who were able to verbalize the concept, as shown in Table 2. In summary, then, it turned out that with instructions that did not specify the nature of the task, age and intelligence were associated with improved achievement in concept attainment on each of three measures of performance.

Experiment II.—Table 3 presents the mean number of errors per group and the results of an analysis of variance of errors, respectively, for Exp. II. In this experiment, in which the nature of the task was explicitly defined, the effect of intelligence disappeared; only age was associated with significant differences in the number of errors. Similar results were obtained from an analysis of the other two measures of performance, number of *S*s who achieved criterion performance and number of *S*s who verbalized the concept, as shown in Table 4.

Combined data from the two experiments.—In view of the comparable variances obtained in both experiments the error data were combined and analyzed for a possible Instructions × *S* interaction. The results of this analysis, presented in Table 5, show that while instructions produced no main effect on performance, the Age × Instruction interaction was significant at the .05 level. When the data were subsequently analyzed separately for each age group, it turned out that only for the 6-yr. *S*s did instructions produce a significant difference in performance ($F = 5.92$, $df = 1.59$, $P < .05$) in favor of explicitness. Another finding, rather surprising, and one difficult to explain, was that the bright 10-yr. *S*s were *handicapped* by the explicit instructions. The mean number of errors for these *S*s under the two conditions was 19.7 and 37.0, respectively ($t = 2.57$, $df = 28$, $P < .02$). The oldest group, the 14-yr. *S*s, performed identically under both conditions. However, as is apparent from their scores, these *S*s performed near the maximum possible level, and, therefore, could not be expected to show improvement under any conditions.

The analysis of the combined frequency data of successful *S*s shows that aside from age, reliable differences were produced by instructions

Table 3: *Mean number of errors per group for reaching criterion of 10 consecutive correct responses with explicit instructions*

Intelligence	GROUPS						Means by IQ
	6-Yr. *Ss*		10-Yr. *Ss*		14-Yr. *Ss*		
	Mean	*SD*	Mean	*SD*	Mean	*SD*	
Average IQ	46.1	25.5	34.2	26.5	14.1	17.4	31.5
High IQ	46.5	34.0	37.0	31.3	7.7	17.4	30.4
Means by Age	46.3		35.6		10.9		

Note.—$F_{Age} = 13.52$, $df = 2$, $P < .001$; $F_{Intel.} < 1$; $F_{Age \times Intel.} < 1$.

Table 4: *Number of Ss who reached criterion and number who verbalized the concept under explicit instructions*

Intelligence	GROUPS						All Ages	
	6-Yr. *Ss*		10-Yr. *Ss*		14-Yr. *Ss*			
	Crit.	Verb.	Crit.	Verb.	Crit.	Verb.	Crit.	Verb.
Average IQ	11	6	12	8	14	12	37	26
High IQ	9	4	9	9	14	14	32	27
Total	20	10	21	17	28	26	69	53

Note.—Criterion data: Age $\chi^2 = 7.081$, $df = 2$, $P < .05$; Intel. $\chi^2 = 1.552$, $df = 1$, ns. Verbalization data: Age $\chi^2 = 17.715$, $df = 2$, $P < .001$; Intelligence $\chi^2 = .046$, $df = 1$, ns.

Table 5: *Analysis of variance of errors for exp. I and II combined*

Source	*df*	*MS*	*F*
Instructions (I)	1	616.05	
Age (A)	2	29,819.62	45.05**
Intelligence (IQ)	1	2,456.81	3.71
I × A	2	2,382.52	3.60*
I × IQ	1	1,798.67	2.72
A × IQ	2	7.90	
I × A × IQ	2	407.87	
Within	168	661.89	
Total	179		

* $p < .05$.
** $p < .001$.

in the Average Intelligence but not in the High Intelligence group. The same results were yielded by the data on correct verbalization.

Rank-order correlations were computed between error scores and MAs in each experiment. The rho values obtained were −.68 under general instructions and −.51 under explicit instructions. Both values are significant at the .001 level. Since the MA is heavily determined by the CA, the value of the correlation per se does not throw light upon the role of the IQ alone. However, as *Ss* in both experiments were identical in age and intelligence, the difference between the two values of rho may be attributed to variation in experimental conditions. Unfortunately, no statistical test exists for evaluating the difference between the rho values; therefore all that can be said is that the direction of the difference is consistent with the other findings.

DISCUSSION

Under general instructions high intelligence was associated with superior achievement in all three measures of performance; but this superiority was not maintained under explicit instructions. If the assumption that specific instructions eliminate the problem-finding component from the task is correct, then the conclusion to be drawn is that high intelligence gave *S* an advantage in problem definition but not in problem solution in the concept attainment task. It is entirely possible, however, that were the concept task one of greater difficulty, intelligence might also prove to be advantageous in the solution phase of the task. This requires further investigation.

Since age was associated with more effective concept attainment on all three measures of performance and under both types of instructions, it may be concluded that maturation involves growth in problem *solving* ability at least until some point between 10 and 14 yr.

In order to assess changes in problem *finding* ability as a function of age, it is necessary to compare performance at each age level under the two types of instructions. That 6-yr. *S*s improved under the explicit instructions is apparent from the significant Age × Instruction interaction in Table 5. The situation was not clear for the 10-yr. *S*s because they actually deteriorated under explicit instructions. As this finding seemed to be inexplicable on any logical basis and was inconsistent with the results pertaining to the other two age groups, it was decided to replicate this portion of Exp. II. Upon replication it turned out that the average 10-yr. *S*s profited from the explicit instructions, while the superior *S*s were unaffected. As to the 14-yr. *S*s, whose performance was the same under both conditions, they did not have much room for improvement, precluding the possibility of reaching a definite conclusion on the effect of instructions at this age.

It thus appears that at least between 6 and 10 yr. *S*s benefited from explicitness of instructions. It is therefore reasonable to conclude that the more effective performance associated with increasing age is not brought about by improvement in problem finding, but rather by more effective problem solving. This points to a possible interesting difference between mental age and intelligence, the effectiveness of the latter being due to more effective problem finding rather than problem solving.

SUMMARY

A concept attainment task was presented under general and explicit instructions to *S*s 6, 10, and 14 yr. of age, divided into two intelligence groups. The results were analyzed in terms of three measures: errors, number of *S*s who achieved the criterion of success, and number of *S*s who verbalized the concept correctly.

Under nonspecific instructions superior intelligence was associated with more effective concept attainment on three measures of performance, but under explicit instructions *S*s of average intelligence improved while those of superior intelligence remained unchanged. Apparently, superior intelligence gave *S*s an advantage in the problem finding phase of the task, but not in actual problem solution.

Performance improved throughout the age range studied. There was no evidence for improved problem finding ability between ages 6 and 10 yr., and it was difficult to evaluate the data for the 14-yr.-olds. It appears, therefore, that problem solving ability, but probably not problem finding ability, increases within the age range studied.

REFERENCES

OSLER, SONIA F., & FIVEL, M. W. Concept attainment: I. The role of age and intelligence in concept attainment by induction. *J. exp. Psychol.*, 1961, *62*, 1–8.

OSLER, SONIA F., & POWELL, M. G. Apparatus for the study of discrimination and concept formation. *Amer. J. Psychol.*, 1960, *73*, 627–629.

42

How Intelligence Tests Limit Understanding of Intelligence

Irving E. Sigel

Sigel's article, in company with Guilford's (3), modifies our understanding of intelligence. Of special note in this paper is the author's concept of "cognitive style," the ways in which an individual organizes his environment conceptually. Sigel is a staff member of the Merrill-Palmer Institute of Behavior and Development, an institute which has served not only as a research center but also as a training agency for workers in the field of child development.

From *Merrill-Palmer Quarterly of Behavior and Development*, 1963, *9*: 39–56

The purposes of this paper are (a) to demonstrate how past and present use of intelligence tests continues to restrict our understanding of intellectual function, and (b) to propose alternatives to the current practice of test analyses. Raising these issues is not to deny the usefulness of intelligence tests—they do provide good bases of predicting academic and vocational success. The contention in this paper is that we are not taking advantage of all the information obtained on an intelligence test.

Intelligence tests usually provide a single summative score, expressed in the form of an IQ, MA, or equivalent. The total score represents the number of correct responses to a set of items of varied content. A unidimensional index is therefore implied in the single score. In reality, intelligence tests are multi-dimensional, not unidimensional, because the items represent a variety of areas of knowledge. The Stanford-Binet is a good illustration of the type of heterogeneous scale yielding a single index.

Test constructors, who are aware of the limitations of single score IQ, attempt to solve the problem by providing subtests, each of which is intended to test particular aspects of intellectual ability. A score is obtained for each subtest. Comparisons of subtest scores enables the examiner to study intra-individual variability of performance. In this way, a more detailed and more sensitive assessment of intellectual ability is possible. An example of this type of test is the Wechsler Intelligence Scale for Children (WISC).

Mindful of these characteristics, let us now turn to a discussion of test limitations. A test response indicates "what" the subject can do, but provides no understanding of "how." To assess these underlying processes, it is necessary to make inferences from the subject's responses. A correct answer to an analogies item, for example, may be a function of particular learning, perceptual discrimination, syllogistic reasoning, or any combination of these. There is no way to judge, from the response itself, which of these processes was operating at the time the response was made.

Not only do we have to infer the underlying process, but also the respondent's repertoire of responses. Only one response is usually required and acceptable. Alternative responses other than those indicated by test manuals are not acceptable. The test constructor decides what is the best answer to items he selects as relevant. The psychometric procedures in test making are too well known to need discussion here. What is not sufficiently taken into account is that, for a number of items, alternative responses—although correct at times—are not considered "correct" or "good" or given equal credit, because of cultural conventions.

The most clear-cut illustrations can be found in verbal tests of analogies. For example, in the WISC similarities item, "How are a scissors and a copper pan alike?" the answer, "They are both household utensils," is given a score of one; the answer, "Both are metal," is given two credits. Both responses are class concepts. Yet, from Wechsler's framework; preference is given to a particular system. How does one decide which system is preferable?

The WISC is not the only test which can be looked at in this way. For example, Guilford (1959)* employed some items to assess abilities to know classes of units. He asks the respondent to "pick one that does not belong" from the following: clam, tree, oven, and rose. Although Guilford does not provide the answer, perhaps assuming the item is simple enough for the psychologists reading the article, the correct answer probably is *oven*. The remaining three items—clam, tree, and rose—belong to a class of living things, but the same items could also be classified on the basis of function—clam, tree, and oven are related to the preparation of food, thus *rose* does not belong; and on the basis of location,—the *clam* may be picked as not belonging, because it lives in water and the other things belong on land. These would probably be wrong answers.

According to Guilford, Wechsler, and other test constructors, classes as units apparently must be organized on the basis of the common hierarchical logical taxonomy. It is a system by which particular characteristics are selected as the basis for a more extensive classification. The characteristics selected or "abstracted" are those "features which enable us to view that thing as an instance or example of indefinitely repeatable patterns or types of situations" (Cohen & Nagel, 1934, p. 371). The features are selected from among an almost infinite variety of properties. Thus, any attribution of similarity between two items (e.g. horse and cow, because they are animals) does not mean that because a horse is an animal it may not be something else. A particular type of class designation has been identified and valued as correct. Respondents deviating from this conception are "wrong" or "inferior."

---

* This article appears in this volume—*Editor's note.*

The conventional class logic is the preferred response.

A word here about "not sharing the conventional class system." The kinds of responses used as illustrations are not bizarre, but bear the reality of consensually agreed attributes of items. Respondents may not share conventional class systems because of different kinds of social and cultural experiences. But non-conventional responses may also reflect the *originality and the novel outlook of respondents*. Aware, as we are, of the obvious cultural and social differences, we still need to be more attentive to differences which emanate from creative thought. Our tests penalize all deviations, regardless of source.

One can disagree with the above contention, because of the assumption that one of the indices of intelligence is the ability to see things as organized in conventional classes of units. If one agrees with this point of view, then intelligence becomes, in part, synonymous with "conformity" in thinking or reasoning.

If this is true, then we are, in effect, saying that a characteristic of intelligence, as measured, is the ability to organize and to classify materials in a conventional way. Is such a notion really part of our conception of intelligence? If it is, should we not make this criterion articulate? . . .

The exhortation to take into account nonintellective factors points directly at personality-type variables. Thus, intelligence test items are used as "projective" tests. To be sure, there is some value in such consideration. On the other hand, we must be ever mindful that such a mode of analysis overlooks "cognitive process" variables which are of equal significance in understanding *how* an individual arrives at an answer. Such concepts as "cognitive style," for example, provide analytic models by which responses can be assessed. Other kinds of cognitive categories can be found in the literature on thinking, problem solving, and reasoning. The plea here is for an evaluation of conceptual-analytic models which will provide standardized scoring systems, taking into account cognitive variables.

A demonstration of how intelligence tests can be used to enlarge our view of intelligence involves discussion of three problems: (a) the relationship of cognitive style variables to performance on intelligence tests, (b) a model for qualitative analysis of verbal content, and (c) what can be learned from analysis of errors on IQ tests.

First, let us turn to a discussion of cognitive style. "Style" is used as an umbrella term, under which is subsumed a variety of cognitive behaviors. It refers to "modes" an individual employs in perceiving, organizing, and labelling various dimensions of the environment.

Operationally, cognitive styles are apparent in everyday life. From a wide array of stimuli, an individual is provided with many alternatives from which to select a basis of organization and labelling. It is assumed that the characteristic selected—especially when little or no coercion is present—emanates from an acquired predisposition to be attentive to that particular type of cue. If given a number of glass tumblers, for example, some individuals will select color, size, quality, or function of the object materials as the most relevant cue by which to organize and subsequently label the stimuli. Since all cues are equally visible and, to a comparable sample of individuals, equally known, we need to explain the fact that different cues are selected and responded to as organizational bases.

Language provides the individual with a set of labels which refer to various aspects of the environment. These labels may refer to parts or to wholes, to attributes, or to the total. The acquisition of the labels is a function of the individual's socio-cultural experiences. The labels acquired enable him to identify and communicate about his environment.

Language, then, is a manifestation of cognitive style. The underlying reasoning is as follows: As an environmental complex, the world of the child or of the adult can be denoted in various terms (i.e., a glass is a drinking object, a cylindrical object, a three-dimensional round object, a useful object, something you use for drinking). Each of these descriptions or identifiers of the object "drinking glass" is accurate but only represents in language some of the attributes of the object under consideration. Many of these kinds of statements about a glass are known to most literate, verbal people. Chances are that if we reversed the procedure, gave a number of these descriptive statements about a glass to different people, and asked them to select the appropriate class term from a list of words among which was the word "glass," the agreement would be very high. In other words, the fact that individuals know a variety of ways of describing an item, does not necessarily enable prediction of how individuals will organize a set of items.

Previous research has demonstrated that individual differences in labelling behavior do exist (Kagan, Moss, & Sigel, 1962). Children and adults, when faced with sorting and, consequently, labelling figures depicting humans and objects, employ labels which can be classified as descriptive, relational-contextual and categorical-inferential labels (Sigel, 1961).

Descriptive. This category includes labels which denote similarity based on manifest objective physical attributes. Subclasses have been identified as, (a) *descriptive part-whole,* labels denoting observable parts of the stimuli; for example, organizing a group of human figures as being similar because "they are all wearing hats" or because "they all have their right arm raised," and (b) *descriptive-global,* labels tending toward denoting the total manifest attributes of the stimuli; for example, identifying a group of uniformed figures as "they are all soldiers" or a group of male figures dressed in identifiable male clothing because "they are men."

Relational-contextual. This category includes labels denoting functional or thematic interdependence between two or more elements in an array. The interdependence is particular to that situation and not generalizable beyond the immediate. The meaning of each item in the grouping is defined in terms of its relation to other items in the grouping; for example, from an array of pictures, a horse and a stagecoach are selected because "a horse is used to pull a stagecoach," or, on a thematic basis a man with a cane who is wearing dark glasses and a young boy are selected because "the boy is helping the blind man to cross the street."

Categorical-inferential. These labels refer to inferred characteristics of the stimuli and each item in a grouping is representative of the total class or category label. Some subclasses of categorical-inferential are, (a) *functional,* where objects are grouped together on the basis of inferred use, (e.g., referring to a group of tools as "these are things to build with"), (b) *class-naming,* where a taxonomic class label is used (e.g., animals, tools, human beings), and (c) *attribute selection,* where only one particular attribute is inferred (e.g., "they are angry," or "they run by motors").

These styles of categorization are operating for the respondent whenever he is faced with the task of organization of items on the basis of similarity or belongingness. The assumption is that individuals have preferences for particular modes of categorization and these characteristic modes have stylistic properties. The tendency will be to classify on the basis of preferred modes whenever the stimuli allow such behavior to manifest itself.

In this framework, let us look at Wechsler's item, "How are a cat and mouse alike?" Since the task requires a perception of similarity, an individual inclined toward a descriptive part-whole approach may say, "They are alike because both have tails." A categorical-inferential response would be, "Both are animals." Note that these two kinds of responses are both correct, even by Wechsler standards. The descriptive part-whole response is, however, valued lower—it gets one point, while the latter response receives two.

On many tests, the scoring of vocabulary items reflects similar problems. The Stanford-Binet test is an excellent illustration. It does *not* take into account the significance of various types of responses (Terman & Merrill, 1960). Only plus and minus scores are given, thus no credit is given for level of response, or for various types of answers. With the word "scorch," for example, credit is given for using a synonym, such as "singe," or an illustration, to "burn your clothes with an iron." It is apparent that these are qualitatively different, but are treated equally. Here again is a type of test in which no system is provided to examine, in an orderly fashion, the quality of the response.

The lack of precise and systematic scoring methods which take advantage of the multiple contributions of a response is but one way in which current use of psychometric tests restrict our view of intelligence. Another restriction lies in the fact that the tests do not provide for adequate assessment of an individual's repertoire of responses. A single response is all that is required. The first response in a vocabulary or a similarities type item may reflect the preferred mode at which an individual responds, but does not tell the level at which he *can* respond. Thus, when an individual says that a cat and mouse are alike because "they have tails," he may be functioning at a level he finds "preferable." This does not mean, however, that he is *unable* to provide a more "abstract" answer if required to do so. There is no provision for such an opportunity.

To this point, we have discussed how the current use of intelligence tests does limit our understanding of the nature of intellectual functioning because of the ways we test and handle test responses. Still before us is one question of whether it would be any better if we did provide for alternative responses, alternative scoring, or analytic procedures. What evidence is there to support the invocation of the concept of style as a significant variable? To this end, we shall discuss some current research.

Scott (1962) assessed the significance of styles of categorization as a predictor of acquisition of science concepts. He found that—for eleven-year-old boys and girls—the better the performance in a task to measure science concept achievement, the greater the use of categorical-inferential responses in sorting tasks. No significant correlation was found between the IQ test and the use of categorical-inferential labels, or between the IQ test and the science achievement test. In other words, the use of categorical-inferential labels on a sorting task was a better predictor of success in the science test than the IQ measure.[1]

Styles of categorization may be an important intellectual dimension to determine how intelligence operates. An individual's "style" dictates the cues he will use, but not necessarily determines the level on which he performs. The style of categorization sets the direction but not the level on which an individual's intelligence might function.

To illustrate further relationships between cognitive style and intelligence, let us turn to a current investigation of the age changes of styles of categorization among "high," "medium," and "low" text-anxious second and third grade boys and girls (Sigel, 1961). The aims of the study are to determine the relationship between styles of categorization and content to be categorized, as well as the relationship between styles of categorization and IQ, age, and test-anxiety level.

Styles of categorization were assessed by using three sorting tasks: (a) arrays of human figures all familiar to the child, (e.g., fireman, woman, boy), (b) object-animal figures, (e.g., items representing vehicles, animals, food), and (c) geometric forms varying in size and color, (e.g., circles, triangles, squares).

To illustrate, we analyzed the relationship between styles of categorization and performance on the Lorge-Thorndike Primary Form. The Lorge-Thorndike is a three-part, non-verbal intelligence test. We used two of the three parts. The subtests employed for the analysis were as follows: (a) a task requiring the child to identify one item out of four that does not belong, and (b) a task requiring the child to find those items, among four, that go together.

Because the child is required to discriminate in both tasks, and determine either an irrelevant item or find similar items, he thus has a requirement similar to that in the IQ test. Consequently, we assume that styles of categorization should be related to his performance on the Lorge-Thorndike.

For the high-anxious third grade boys and girls, the use of class-naming labels—a type of categorical-inferential categorization—was related significantly with the total IQ score. For these same children, the more select descriptive-global responses made, the less "able" they were in selecting the item that did not belong in a series of items. Thus, those children who organize material on the basis of global manifest characteristics made errors on tasks requiring perception of structural similarities. In other words, their performance on this test is related to the styles of categorization employed by these children. The success with which children perform on IQ tests is, to some measure, predictable from their styles of categorization. These findings, thus far limited to high test-anxious children, tend to support our contention that performance on an intelligence test measure *is* influenced by a style of categorization.

Another finding relevant to intelligence test performance is that a positive relationship exists between the frequency of use of descriptive part-whole responses and cautiousness. There is some indication that such a preference for categorization tends to call forth a reflective careful approach in making judgments. In fact, Kagan (1962) concludes that reflectiveness is associated with a preference for a descriptive part-whole approach. The concept of reflectiveness or cautiousness as used here can be considered a cognitive set, or approach, and certainly would be manifest in test performance.

How extensive the role of styles of categorization is for all the classes of items employed in intelligence tests is an open question. What can be said at this point is that there is enough evidence from a deductive-analytic basis, as well as an empirical one, to suggest that IQ test scores are influenced by the style of categorization of the individual. The larger task is to determine, with greater specificity, the degree to which such relationships exist, and whether particular styles facilitate or inhibit intellective responses.

Taking this position calls for new ways of scoring our tests, as well as new ways of constructing them. Tests can no longer be scored by the test constructor's convention, nor should there be provision for only one type of response. We shall need to re-examine our testing materials, especially the selection and scoring of items, if we are to assess the relevance of the arguments proposed here.

---

[1] The IQ measure used was the Southend Test of Intelligence, London, England. George Harrup, publisher, 1953.

Styles of categorization provide but *one* approach to the study of IQ responses. Other models for qualitative analysis could be devised. An illustration of this is a study reported by Feifel and Lorge (1950), in which they examined Stanford-Binet vocabulary records to ascertain qualitative changes with age in concept formation. As we shall see, the results of this study show quite dramatically what can be gained from additional scoring systems.

Feifel and Lorge were interested in the successive stages of concept formation, and development of thinking. They studied verbatim reports of nine hundred school children, boys and girls, ranging in age from six to 14. The categories employed were a synonym category, illustration and inferior explanation, and an error category. Their results (1950) were as follows:

> Younger children significantly more often employed use and the description category and illustration demonstration, inferior explanation, and repetition of response; whereas older children significantly used synonyms and explanation type responses (p. 2).

The Feifel and Lorge study provides one kind of framework for setting expectations regarding the age changes in thought. Within this framework, norms can be created which would indicate the child's level of conceptual thought, not possible even from the 1960 revision of the Stanford-Binet. In this way, two types of information can be obtained, how the child thinks and what he knows.

What is of considerable interest, is that current revisions of the Stanford-Binet do not take advantage of what the test really has to offer in the way of understanding intellectual functioning. It certainly would have been a considerable step forward if the test constructors had made substantive improvements incorporating some of the fine studies done with this test. In the long run, we would have had better indices of intellectual performance.

One last point on cognitive styles, although tangential, is of cardinal importance. In all our work on styles, a most consistent finding was the phenomena of "sex differences." Not only do boys and girls differ in their styles of categorization, but more important is that the psychological role played by particular styles differs for boys and girls. Personality correlates of styles of categorization differ for the sexes. We find, for example, that the descriptive part-whole approach relates positively with cautiousness for boys, but negatively for girls. The sex difference seems to be basic. Yet, in our intelligence test scores, we do not have separate norms for boys and girls. If we extrapolate from the research on style of categorization, it is reasonable to suppose that sex differences should show up in a host of intellectual operations.

Two kinds of evidence have now been presented to support the contention that intelligence tests, as now viewed, are restrictive. The suggestion has been made persistently that additional scoring systems are necessary in order to extract the meat from the responses, rather than our current approach which is so parsimonious that we lose perhaps more than we gain.

Another area to examine is how the analysis of errors on IQ tests can provide us with useful information. This argument derives from disagreement with the focus on "correct" responses. The assumption is that errors are not random or accidental. They are manifestations of a quality of intellective operations, comparable to the correct response. This is especially the case in those tests where choices are provided, each of which is purposeful, and taps some kind of function.

Errors have been analyzed unsystematically, by clinically oriented examiners, to arrive at a more comprehensive understanding of the particular case. For example, reference is frequently made in reports dealing with the quality of the responses, about the degree to which responses are within the realm of the expected or not. Clinical psychologists, influenced by the Rapaport approach, have been examining Wechsler-Bellevue items from a projective point of view. Thus, errors in the picture arrangement, for example, have been analyzed in personality terms, presumably reflecting intra- or interpersonal conflict.

Such analyses are subject to considerable error because they are not based on a systematic analysis using standardized schema within a comparable conceptual framework. It seems that this is a necessary condition to allow for the exploitation of data in a way that would provide wide comparability. The tragedy is that much data are being overlooked which, if properly used, would add to our knowledge of intellectual functioning.

If the assumption is correct that errors are determined by particular cognitive operations, a systematic analysis of errors should reveal facets of intellectual operations not revealed in the total score of the test. A test which provides such an opportunity is the Raven Progressive Matrices (PM), (Raven, 1938), a non-verbal perceptual-type test made up of a number of designs or matrices, each of which has a piece missing. The respondent examines the matrix and, from a number of choices at the bottom of the page, selects the appropriate one to complete the pattern.

The PM, according to Raven, is a test of "innate educative" ability. It is a test which is supposed to assess the individual's capacity to form comparisons, reason by analogy, and develop logical methods of thinking, regardless of previously acquired information.

If the Raven PM is supposed to measure the intellectual qualities described, then the kinds of errors made should reveal the limitations of the respondents. Further, since the respondents are required to make choices, the kinds of choices made should reveal the quality of intellectual performance. Thus, if variety exists in the choices allowed, only through analysis of all types of choices made can one discover the unique patterns of functioning. If one answers all items correctly, we learn something different about him.

Raven was apparently interested in error analysis; he provides a set of categories in his test manual by which errors can be analyzed (Raven, 1956). No one has systematically done this with large populations of "normal" individuals. Using Raven's system, we analyzed 1,000 records of boys and girls ranging in age from nine to 18. In another study, we compared types of errors of seven-, nine-, and 11-year-old boys to study the positional responses as a source of error.

Each of the sets in the PM are defined as assessing a distinctive process. Set A is described by Raven (1956) as a "simple continuous pattern—completion involving perception of difference, similarity and identity of pattern only" (p. 30). Each choice item reflects some aspect of these factors. The choices as described by Raven (1956, p. 30) are:

Choice 1. The figure is wrongly oriented and is incomplete as far as it goes. This is classified as incomplete correlate.

Choice 2. This figure has similar properties to Choice 1 except that the part corresponding to the correct piece is horizontal.

Choice 3. The figure shown is irrelevant and is classified as a different response.

Choice 4. This figure is contaminated by irrelevancies or distortions and is classified as *inadequate individuation.*

Choice 5. This choice is an instance of *incomplete correlate,* (i.e., figure wrongly oriented).

Choice 6. This is the correct choice and completes the matrix.

With these categories in mind, examination of errors, item by item, should provide us with the specificity of the child's functioning not obtained when just tallying the number of correct responses and referring only to the percentile score. Since it is the rare child who is able to do all the problems correctly, examination of the *kind* of error should provide valuable diagnostic data. In this presentation, we shall focus on some error analyses in order to point up what additional information can be gleaned from such an approach.

To illustrate, look at the responses made to PM Item A-7 (Figure 1). In our analysis, comparisons between boys and girls, as well as two age groups, will be presented to demonstrate the significance of sex as well as age differences.

Table 1 provides the percentage of choices made by nine- and ten-year-old boys and girls. Boys and girls within and between age groups do not differ significantly in the percentage of correct choice. The differences appear in the kinds of errors made. Nine-year-old boys and girls, in erring, selected Choice 4 with greatest frequency. Boys, however, make this error considerably less than the girls.

Girls, in other words, make more of their errors because of difficulty in perceptual discrimination. This finding suggests that more girls at this age are global in their approach than the boys. . . .

Before discussion of the implications of these findings, let us examine one more PM item (Raven, 1956) calling for different intellectual abilities, one from Set B. Set B is presumed to measure "apprehension of analogous changes in spatially and logically related figure" (p. 32). According to Raven, Item B-7 (Figure 2) is intended to measure "concrete or coherent reasoning by analogy" (p. 32). The choices are: *Choice 1.* Repetition of figure. *Choice 2.* Repetition of figure. *Choice 3.* Inadequate individuation. *Choice 4.* Incomplete correlate. *Choice 5.* Is correct. *Choice 6.* Incomplete correlate (Raven, 1956).

Table 2 indicates the kinds of choices made by nine- and ten-year-old boys and girls on item B-7. Inspection of the table reveals that more nine- and ten-year-old boys selected the correct item than girls of similar ages.

When we examine the kinds of errors nine-year-old children make, we find that girls more frequently select Choice 1, a repetition of the

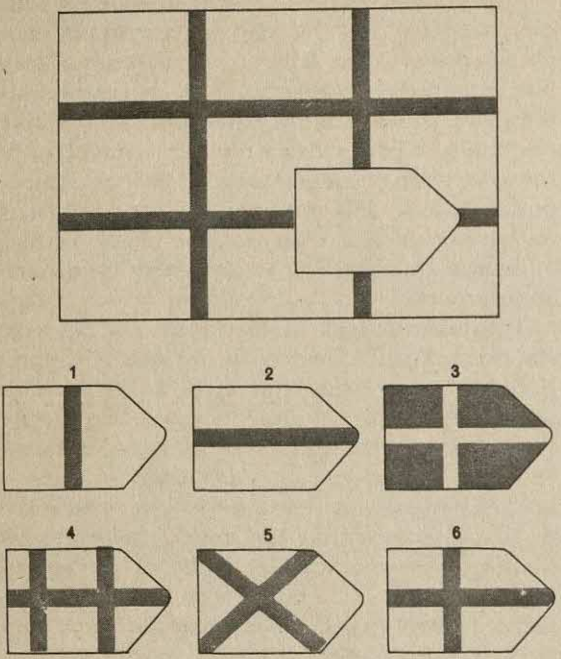


Fig. 1: *Item A-7 from Raven, J. C. Progressive Matrices (H. K. Lewis & Co., Ltd., London). Reprinted with permission.*

figure. The boys select Choice 2, which is also classified by Raven as a repetition of the figure. It is interesting that boys and girls make the same type of error but select a different figure. The difference may be because the girls, in reading the pattern from left to right, perseverate and select Choice 1. Boys may perseverate vertically and select Choice 2. If we combine Choices 1 and 2, since both are errors of "repetition of the figure," we find that this type of error is comparable for the nine-year-olds (35.7 per cent of the girls and 35.3 per cent for the boys). At age ten, more children get the correct item, but when errors are made, repetition of figures choices are again the most frequent (43.3 per cent and 40.9 per cent respectively for girls and boys, when Choices 1 and 2 are combined). Other changes occur, somewhat more pronounced for girls in this case. There is a drop in errors of the inadequate individuation kind. The implications of this datum in relation to the discussion of item A-7 need further consideration. We can conclude, however, that for children erring, the perseverative choice increases a year later. It suggests that this factor may be an important symptom of why the children make errors. This occurs in spite of the increases in the number of children making correct responses.

The detailed presentation of item analyses from the PM are illustrative of the kind of information to be gained from examination of errors. For Item A-7, we found that girls tended to have difficulty in perceptual discrimination, while for Item B-7 boys and girls tended to make perseverative errors. In both instances, the trends are similar at age nine and ten for each sex group. Although the age and sex differences as reported here are not statistically significant, they seem to be of sufficient magnitude to warrant further study. Real age and sex differences might be sharpened by conducting experimental studies in which the variables discussed are controlled with a precision not attainable in a test constructed for other purposes.

This kind of information would not be available were we to limit our analysis to the correct response only. Using this procedure for groups or for individuals provides us with information on *how* children are responding to an item. Thus, we may discover the degree to which the child is functioning maturely or not. For example, examination of items like B-7 enables us to assess the degree to which groups or individuals perseverate. Perseveration may be a function of emotional factors, (e.g., anxiety). It could also be interpreted from a cognitive frame of reference. Spe-

Table 1: *Percentage of choices made by two age groups of boys and girls on PM item A-7*

Choice	*Girls* 9 yrs. (N = 42)	*Boys* 9 yrs. (N = 34)	*Girls* 10 yrs. (N = 53)	*Boys* 10 yrs. (N = 49)
1	0	0	3.7	2.4
2	4.8	8.9	1.9	6.1
3	4.8	2.9	3.7	2.0
4	21.4	11.8	20.7	10.2
5	0	0	0	0
6	69.0	76.4	69.8	79.3

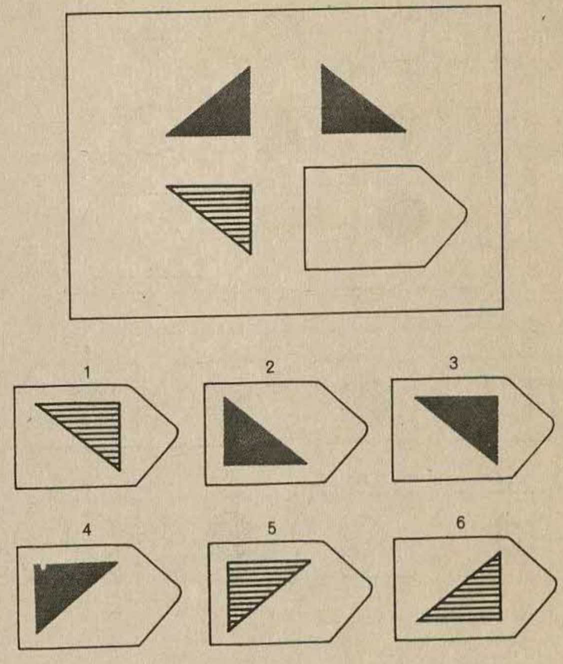
cifically, from the latter viewpoint, perseveration is an inability to shift one's own framework, and reflects difficulty in judging each choice as a separate. The error is a manifestation of the child's stimulus-boundness and his difficulty in reorienting himself. Of course, the test does not and cannot provide antecedent explanations of these performances. However, the test can and does provide a more comprehensive picture of how children function intellectually in the here and now.

It might be argued that the total score would indicate the degree to which the child is functioning "maturely" or not. Inspection of our data reveals that although children may be consistent within themselves in the errors they make, there seems to be no relationship between type of error and total score. For example, children who err on B-7 by using a perseverative choice (generally considered a less mature response) range from very high to very low total scores on Set B.

In the second study with the PM, the role of position and perseveration of items was studied as a factor in errors. Records of three groups of boys, ages seven, nine, and 11, were examined. The matrix was divided into proximal and distal sections. The proximal field included designs items surrounding the missing piece. The distal field included designs on the periphery. Among the choices are items identical to those in the proximal and distal fields. Figure 3 illustrates the division of the matrix.

The younger children, being more stimulus bound and more perseverative, would be expected, in erring, to select items identical with the matrix items in the proximal field; the older children, to select items identical with the matrix items in the distal field. It was found that significantly more seven-year-olds selected, when erring, the proximal item. The nine-year-olds split about evenly between the proximal and distal choices, and the eleven-year-olds showed a significantly greater selection of the distal choices. We can conclude that errors of younger children are a function of the "principle of proximity," in effect substantiating our predictions.

Fig. 2: *Item B-7 for Raven, J. C. Progressive Matrices (H. K. Lewis & Co., Ltd., London). Reprinted with permission.*

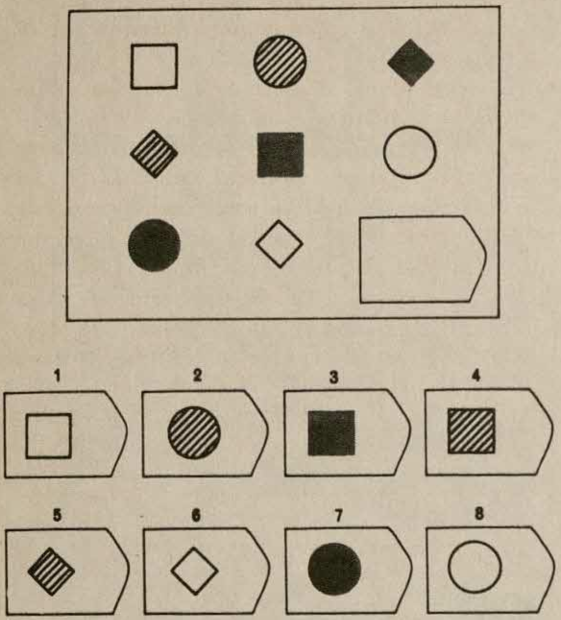


Not only did the younger children select the items in the proximal field, but, in doing so, tended to select the one identical to the item preceding the missing one. Whether or not the potency of a tendency to perseverate accounts for the high incidence of proximal choices is a question. There are indications that when an identical choice is occasionally available distally, a distal choice is made by seven-year-olds. There

Table 2: *Percentage of choices made by two age groups of boys and girls on PM item B-7*

Choice	Girls 9 yrs. (N = 42)	Boys 9 yrs. (N = 34)	Girls 10 yrs. (N = 53)	Boys 10 yrs. (N = 49)
1	28.6	14.7	26.4	22.5
2	7.1	17.6	16.9	18.4
3	4.8	2.9	0	4.1
4	7.1	0	0	2.0
5	38.1	50.1	47.3	63.3
6	14.3	14.7	9.4	16.3

Fig. 3: *Illustration of type of item used to assess proximal or distal choices of young children.*



are not sufficient items in the test to determine whether identity leads to proximal choices or proximal choices occur because of identity.

In any event, the responses made are functions of more than the child's knowledge. If, as Piaget holds (Inhelder & Piaget, 1958), children of this age are concretistic and do not utilize logical operations, the errors may reflect the quality of intellectual functions. Ironically, we do not know what may account for the correct response, but analysis of errors could tell us what kinds of difficulty the child is having.

The type of error analysis discussed provides us with increased specificity of the quality and level of intellectual performance not possible from a consideration of correct responses only. To be sure, the number of correct responses presumably provides us with a measure of what the respondent can do. Analysis of errors should tell us more about the "why not's" of performance.

Another contribution of error analysis, accepting the diagnostic significance of them, is for our understanding of development. Developmental curves based on errors could reveal how children alter their approach with increase in age. To do this would require a conceptualization of errors so that an "erring" response can be identified as having particular significance. The PM provides such a model. No doubt the same could be done with other tests.

Although we have not related errors on the Raven PM to styles of categorization, it may well be that particular styles would be inhibitors to correct responses. For example, the child who tends to be global in his approach may err on the Raven PM item requiring perceptual discrimination, but not on the item requiring other cognitive operations.

Thus far, the focus has been on styles of categorization, qualitative analysis of particular test responses, and, finally, errors. The "style" issue has provided one factor that may account for the kind of response children give. The distribution of qualitative analysis and errors provide methods to increase the information already available on the test as well as to enhance understanding of the individual's intellectual behavior.

Why should new tests not be developed rather than spend time revising the old? Aside from the difficulty of constructing a test, there is the question of how ready are we? It seems we are at the beginning of a period which should enable the creative test constructor to integrate various elements of knowledge regarding cognitive functioning. We must, however, wait until the scientific value of current research in cognitive functioning is established. The concept of cognitive styles is a good case in point. To date the evidence is convincing that individuals do tend to approach classes of material in a consistent manner. If this continues to hold, and if we can show the predictive value of such phenomena, then it is conceivable a test of intelligence could be constructed to tap these dimensions. An item like this could be used: "Rank an item that best goes with cow," and the choices are animal, meat, tail, feet, shoes. Each of these is correct, yet each reflects a different category of response. Perhaps in this way we can get at ability and style.

Resting with our current concepts will not help us understand the "how" of intelligence. Using intelligence tests as we do, limiting our analysis to unsystematic analyses of responses, be they qualitative or quantitative, negates the relevance of other cognitive factors influencing IQ. It may well be that the reason IQ and creativity are not frequently related is just because we have used the IQ in a narrow restricted way. . . . If we broaden our use of IQ tests, we may be in a position to provide a significant test which would not only have broad predictive value, but enlarge our view of how intelligence operates.

In this way, we may avoid such curious contradictions that IQ and creativity are not related. Such findings are essentially artifacts of our test construction. The greater tests reflect theoretical conceptions, the greater our chance of clarifying many of the issues plaguing us.

REFERENCES

COHEN, M. R., & NAGEL, E. *Logic and scientific method.* New York: Harcourt, Brace and Co., 1934.

FEIFEL, H., & LORGE, I. Qualitative differences in the vocabulary responses of children. *J. educ. Psychol.*, 1950, *41*, 1–18.

GUILFORD, J. P. Three faces of intellect. *Amer. Psychologist*, 1959, *14*, 569–579.

INHELDER, B., & PIAGET, J. *The growth of logical thinking from childhood to adolescence.* New York: Basic Books, 1958.

KAGEN, J., MOSS, H. A., & SIGEL, I. E. The psychological significance of styles of conceptualization. In J. F. Wright & J. Kagan (Eds.), Proceedings of conference on cognitive processes. *Monogr. Soc. Res. Child. Develpm.*, 1962.

RAVEN, J. C. *Progressive matrices.* London: H. K. Lewis, 1938.

RAVEN, J. C. *Guide to using the coloured progressive matrices.* London: H. K. Lewis, 1956.

SCOTT, N. The relationship of inductive reasoning and cognitive styles and categorization behavior to science concept achievement in elementary school children. Unpublished doctoral dissertation, Wayne State University, 1962.

SIGEL, I. E. *Cognitive style and personality dynamics.* Interim progress report for National Institute of Mental Health, M-2983, 1961.

TERMAN, L. M., & MERRILL, MAUD A. *Stanford-Binet intelligence scale.* Boston: Houghton Mifflin, 1960.

PERSONALITY DEVELOPMENT

43

Relationship Between Level of Aspiration and Selected Physical Factors of Boys Aged Nine Years

H. Harrison Clarke and David H. Clarke

The preadolescent years bring increasing awareness of one's body; this is especially true for boys because of the peer emphasis on skill. The Clarkes' study demonstrates the interrelationship between physical adequacy and self-esteem. The concept of "level of aspiration" has been used in many studies and owes its origin to the work of Kurt Lewin.

A technique of individual personality evaluation seems to be offered in the use of level of aspiration testing. . . . The concept of approaching certain environmental, personal, and social factors through the use of this procedure has suggested possibilities for improving understanding of human motivation, one of the difficult types of psychological assessment.

REVIEW OF LITERATURE

Only one published level of aspiration study has been related specifically to physical education. Smith (1949) investigated the influence of

---

From *The Research Quarterly*, 1961, *32*: 12–19. Reprinted with permission of the American Association for Health, Physical Education, and Recreation.

football success and failure on the level of aspiration of the participants. His findings agreed with those of Festinger (1942), who worked with other types of activities, in that there was a tendency for successful players to raise their level of aspiration. Further, Smith found that individuals expressing higher levels of aspiration maintained some hope for success while repeatedly experiencing failure, and that those with low aspirations who experienced failure overtly withdrew from failure-producing situations. He also observed a tendency for those with the highest levels of aspiration to experience success repeatedly, even though they raised their level of aspiration goals. Smith concluded that the ultimate level of aspiration was not correlated with ultimate achievement.

Inasmuch as no study has investigated level of aspiration utilizing a measure of maximum strength or has examined its relationship to selected physical, motor, and maturity factors, such an investigation seems appropriate. This research was undertaken to shed additional light on the role played by physical factors when associated with the achievement of aspiration goals.

METHODOLOGY

Level of Aspiration Test

Level of aspiration, as defined by Frank (1935), is "the level of performance on a familiar task which an individual undertakes to reach." It was determined in the present study with the use of a hand dynamometer designed to measure the maximum static strength of the hand-gripping muscles. Each subject was given this test, told of his results, and asked to predict what his next effort would produce. He was then given a second grip test. On the basis of his second performance, the subject was again asked to assess his next grip test. He was then given a final strength measurement. Two scores only were utilized in this study to assess level of aspiration, as follows:

1. *Discrepancy Score 1 (DS-1).* This score was assessed in accordance with the technique used by Boyd (1952), which consisted of determining the difference between the first grip strength score and the first stated aspiration score. A positive value was recorded when the aspiration score was greater than the subject's preceding grip strength and a negative value was given when it was less.

2. *Discrepancy Score 2 (DS-2).* This discrepancy score was computed in the same manner as DS-1, except that each subject's second grip strength achievement and his second aspiration expression were utilized.

Experimental Variables

Maturity Assessment. Skeletal age was determined by assessing the X ray of the hand and wrist against the Greulich and Pyle standards (1950).

Structural Measures. Body weight, standing height, flexed-tensed upper arm girth, hip width, McCloy's Classification Index I (1939), and Wetzel Nutritional Channel (1943).

Motor Test. Standing broad jump.

Strength Tests. Rogers' arm strength, McCloy's arm strength, Rogers' Strength Index, Rogers' Physical Fitness Index, and the mean of the following 12 cable-tension strength tests: elbow flexion, shoulder flexion, shoulder inward rotation, trunk flexion, trunk extension, hip flexion, hip extension, hip inward rotation, knee flexion, knee extension, ankle plantar flexion, and ankle dorsal flexion.

Subjects

The subjects were 98 nine-year-old boys in the Medford, Oregon, public schools. The tests were administered within two months of each boy's birthday.

Testers

Before collecting research data, the testers administered the anthropometric tests (upper arm girth and hip width), the standing broad jump, and all cable-tension strength tests twice to 30 subjects. The results of the two administrations of each test were then correlated to obtain objectivity coefficients. If the resultant correlation did not compare favorably with that established for a test, the testers continued to practice until a satisfactory degree of agreement between testers was realized. The correlation coefficients achieved for the cable-tension strength tests ranged from .87 to .99, which approximated those reported for these tests; those for the anthropometric measurements were in the high 90's. The tests comprising the battery for the Strength Index and Physical Fitness Index were administered by testers with extensive testing experience.

Treatment of Data

For each of the discrepancy scores (DS-1 and DS-2), the subjects were divided into three groups: those with the highest positive scores; those with zero, or no discrepancies; and those with the greatest negative scores. These were distinct groups with gaps in the distributions between them. Of the 98 boys included in the study, 48 and 54 were used in the three DS-1 and DS-2 groups respectively. The characteristics of these distributions of discrepancy scores were as follows:

	DS-1			DS-2		
	N	Mdn	Q	N	Mdn	Q
High Positive (High) Group	20	8.00	2.00	20	6.00	2.00
Zero Group	14	—	—	22	—	—
High Negative (Low) Group	14	−2.50	1.50	12	−3.50	4.50

The differences between the means of the high, zero, and low aspiration groups for the various physical factors were computed separately. These differences were tested for significance by application of the *t* ratio applied to uncorrelated groups. Significant *t* ratios at the .05 level ranged from 2.02 to 2.06 for the various degrees of freedom; at the .01 level, the range of significant *t* ratios was 2.70 to 2.78.

RESULTS

Discrepancy Score 1. . . . A summary of . . . results follows.

1. *High vs. Low Discrepancy Groups.* The boys with the highest DS-1's had significantly greater mean Physical Fitness Indexes (115.60) than those with lowest DS-1's (100.30). The difference between these means was 15.30. This difference is significant at the .05 level, since the *t* ratio is 2.21. None of the differences between the means for the other experimental variables was significant.

2. *High vs. Zero Discrepancy Groups.* When divided on the basis of high and zero DS-1's, several experimental variables exhibited differences between means. The differences between the means for the Physical Fitness Index and McCloy's arm strength score were significant at the .01 level; the *t* ratios were 2.90 and 2.82 respectively. The following four additional tests had differences between means which were significant at the .05 level and above: height, weight, McCloy's Classification Index, and Rogers' Strength Index.

3. *Low vs. Zero Discrepancy Groups.* None of the *t* ratios representing the difference between means for the various experimental tests was statistically significant when the low and zero DS-1 groups were compared.

Discrepancy Score 2. . . . None of the *t* ratios for either high vs. low, high vs. zero, or low vs. zero DS-2 groups was significant.

SUMMARY AND CONCLUSIONS

Two assessments of level of aspiration were made with the use of a hand dynamometer, and 13 physical, motor, structural, and strength tests were given to 98 nine-year-old boys. Two aspiration discrepancy scores were computed and the subjects were divided into three extreme groups: those with the highest positive scores, those with zero discrepancies, and those with the greatest negative scores. Mean differences for the various physical factors among these groups were computed and tested for significance. The following results were obtained:

1. The Physical Fitness Index, a test of muscular strength relative to each boy's age and weight, was the only test in which the differences between means were significant in more than one comparison of the aspiration groups. The boys in the high DS-1 group had a significantly higher Physical Fitness Index mean than did the boys in the zero and low DS-1 groups. However, significant differences between the means for this test were not obtained when the low and zero DS-1 groups were compared; also, none of the differences between the Physical Fitness Index means was significant for the comparison of the DS-2 groups.

2. The greatest number of significant differences between means on the various experimental tests was obtained when the high and zero DS-1 groups were compared. The differences on the following six tests were significant at the .05 level and above: height, weight, Classification Index, McCloy's Arm Strength, Strength Index, and Physical Fitness Index.

3. None of the differences between the means of the low and zero DS-1 groups and between the means of all DS-2 groups was significant. Further, the following tests were not significant for any comparison between level of aspiration groups: upper arm girth, hip width, Wetzel Nutritional Channel, Rogers' Arm Strength, cable-tension strength test average, standing broad jump, and skeletal age.

Apparently, the nine-year-old boy who strives to attain higher goals (expresses higher levels of aspiration) is physically superior in size and strength to others his age who are not willing to risk the chance of failure, and who thereby choose the aspiration level that seems to ensure at least some measure of continued success. It would seem that this technique does provide a means for studying the behavior of young boys, for their selection of level of aspiration appears to reflect the previous success or failure which they associate with the task. This has been represented in the present study by the grip strength test.

REFERENCES

BOYD, G. F. The levels of aspiration of white and negro children in a non-segregated elementary school. *J. soc. Psychol.*, 1952, *36*, 191–196.

FESTINGER, L. Wish, expectation, and group standards as factors influencing level of aspiration. *J. abnorm. soc. Psychol.*, 1942, *37*, 184–200.

FRANK, J. D. A comparison between certain properties of the level of aspiration and random guessing. *J. Psychol.*, 1937, *3*, 43–62.

FRANK, J. D. Individual differences in certain aspects of level of aspiration. *Amer. J. Psychol.*, 1935, *47*, 119–128.

GREULICH, W. W., & PYLE, S. I. *Radiographic atlas of skeletal development of the hand and wrist.* Stanford: Stanford University Press, 1950.

MCCLOY, C. H. *Tests and measurements in health and physical education.* New York: F. S. Crofts and Co., 1939.

SMITH, C. H. Influence of athletic success and failure on the level of aspiration. *Res. Quart.*, 1949, *20*, 196–208.

WETZEL, N. C. Assessing the physical condition of children. *J. Pediatrics*, 1943, *22*, 82–110, 208–225, 329–361.

44

The Relationship Between "Causal" Orientation, Anxiety, and Insecurity in Elementary School Children

Rolf E. Muuss

Muuss supplies more details on the Preventive Psychiatry Program of Ojemann (33). In this article the effects of the planned learning program in developing the "causal" orientation in school children are examined in relation to emotional adjustment. Both observation and self-report are used as assessment techniques with the experimental and control groups.

This paper presents the findings of a study designed to investigate the relationship between "causality" and such indices of mental health as anxiety and insecurity in fifth- and sixth-grade children. The concept, causality, constitutes the theoretical framework of the Preventive Psychiatry Program at the State University of Iowa. The Preventive Psychiatry Program is designed to investigate the extent to which causal orientation contributes to mental health, and whether and to what extent an experimental learning program emphasizing the causal nature of human behavior produces causally oriented subjects.

Causality is defined as an understanding and appreciation of the dynamic, complex, and interacting nature of the forces that operate in human behavior. It involves an attitude of flexibility, of seeing things from the view point of others as well as an awareness of the probabilistic nature of knowledge. A causally oriented person is capable of suspending judgment until sufficient factual information is available; furthermore, he realizes that his behavior has consequences and that there are alternative ways of solving social problems. It is assumed that a person who is aware of the dynamic and causal nature of human behavior is better able to solve his own problems and to meet social situations. This study purports to investigate a limited aspect of this assumption.

Explicitly stated, our hypotheses are that *S*s

From *Journal of Educational Psychology*, 1960, *51*: 122–129

who are high causally oriented (as measured by paper-and-pencil tests) will differ from *S*s who show a low degree of causal orientation on the following criterion variables:

Hypothesis 1. They will demonstrate less insecurity as measured by the Kooker Security-Insecurity Scale (Kooker, 1951).

Hypothesis 2. They will show less anxiety as measured by the Children's Manifest Anxiety Scale (CMAS) (Castaneda, McCandless, & Palermo, 1956).

Hypothesis 3. They will make fewer L responses as measured by the L scale which is part of the CMAS.

The rationale for these hypotheses is based on the assumption that a lack of insight into the dynamics of one's own behavior and an unwillingness and/or an inability to understand the problems and the behavior of others tends to increase the level of anxiety and the degree of insecurity. If other people's behavior is not understood it will tend to be threatening, as are physical events which an individual experiences and does not understand. A lack of insight into the dynamics of behavior will tend to make it difficult to react logically to the behavior of others. Furthermore, if behavior is not understood, it may be misinterpreted and the individual may react in such a way as to produce a threat to the other person's security and self-respect. This then might tend to generate conflict which would add further to the difficulty of the situation. If at times a person does not understand his own behavior and the factors that influence him he naturally feels threatened, insecure, and anxious. Once a person does understand himself and others he is more willing to agree with such statements as: he sometimes gets angry, he does not always tell the truth, he does not like everyone, and similar items that make up the L scale.

Hypothesis 4. Furthermore, it is hypothesized that these differences—if they have any generality at all—will be equally obvious for fifth- and sixth-grade *S*s.

Hypothesis 5. The two measures used, insecurity scores and anxiety scores, will show a positive relationship to one another.

Hypothesis 6. Since about half of the *S*s have been exposed to an experimental learning program and half have served as controls in another research study, it is hypothesized that a significantly larger portion of the high causally oriented *S*s came from the experimental classes and that a larger portion of the low causally oriented *S*s came from control classes.

PROCEDURE

In order to investigate the above stated hypotheses, two sets of tests were administered to 280 sixth- and 179 fifth-grade *S*s in the schools of a midwestern community of 80,000. Two hundred twenty-four of these *S*s had been exposed to an experimental learning program designed to develop a causal orientation, while 235 *S*s came from regular classrooms and served as controls in another research study. The fact that the *S*s came from both experimental and control classes was only utilized in testing the sixth of the previously stated hypotheses. However, there is some justification in utilizing *S*s from two different groups since there is evidence (Stiles, 1950) that children from regular classrooms do not have much of an understanding and appreciation of the causal nature of human behavior. Therefore, in order to obtain a wide range of causal orientation, the experimental and the control *S*s were combined so that differences between high and low causally oriented *S*s could be studied more effectively.

The following two tests served as selection criteria to determine high causally oriented and low causally oriented *S*s:

1. The Social Causal Test[1]
2. The Physical Causal Test[2]

Of these two tests only the Physical Causal Test had a significant correlation ($r = .58$, N 251) with IQ. This might explain that the high causal group differed significantly from the low causal group in terms of IQ scores. However, this is not a serious shortcoming since it will be shown that the criterion variables, insecurity, anxiety, and L scores, have a relatively low relationship to IQ. The high causally oriented *S*s, for the purpose of this study, are defined as all those *S*s who fell above the corresponding grade mean on *both* the above described selection criteria, low causally oriented *S*s as those *S*s who fell below the corresponding group mean on *both* selection criteria.

The second set of tests, constituting the criterion variables, consisted of:

1. The Kooker Security-Insecurity Scale. Kooker ratings are obtained by a trained observer who follows the *S*s during the whole school day for a period of five days, and who

---

[1] The Social Causal Test is described elsewhere in the literature (Lyle & Levitt, 1955; Ojemann, Levitt, Lyle, & Whiteside, 1955).

[2] The Physical Causal Test was developed in part by the author. Part of the test was modified by the author for use with fifth- and sixth-grade *S*s from Clark (1953).

Table 1: *Distribution of original and corrected number of cases in the factorial analysis of variance*

Grade Level	ORIGINAL N FOR THE SECURITY SCORES		ORIGINAL N FOR THE ANXIETY AND L SCORES		CORRECTED[a] N IN ORDER TO OBTAIN PROPORTIONALITY	
	Degree of Causality		Degree of Causality		Degree of Causality	
	High	Low	High	Low	High	Low
5th grade	59	41	58	45	60	45
6th grade	89	72	89	71	88	66

[a] *This corrected distribution was used for all three sets of scores: anxiety, security, and L scores (see text).*

rates the child on a series of 19 behavior items indicative of security or insecurity as to frequency of occurrence. When several observers independently rated the same *S*s, between rater correlations ranged from .63 to .86 (Kooker, 1951).

2. The Children's Manifest Anxiety Scale (CMAS) (Castaneda, McCandless, & Palermo, 1956).

3. The 11 item L scale which is administered interspersed with the CMAS.

THE DESIGN OF THE STUDY

From the 280 sixth-grade *S*s 90 met the selection criteria and were classified in the high causal group; 72 fell below both group means on the selection criteria and were classified as the low causal group. Similarly, from the 179 fifth-grade *S*s, 59 met the criteria for the high causal group, and 45 fell to the category low causal group. In order to test for the effect of grade level (G) (fifth or sixth grades) and the degree of causality (C) (High and Low as defined above) an analysis of variance was computed for each of the three criterion variables, insecurity, anxiety, and L scores. A factorial analysis of variance design was utilized for all three sets of scores. Since this design requires either an equal number of cases in each cell or proportionality of cases from column to column or from row to row, the data had to be adjusted. Proportionality could be obtained with the least loss of cases by utilizing the corrected distribution shown in Table 1. Proportionality of cases—as illustrated in Table 1 under Corrected N—was obtained by either randomly omitting cases from the original N, or by adding cases made up of the corresponding cell group mean of the original N. The original N's for the security and the anxiety scores are also presented in Table 1.

RESULTS

The analysis of variance data are reported in Table 2. Inspection of Table 2 reveals that there is a highly significant ($p < .001$) difference in the predicted direction between the high causally oriented and the low causally oriented *S*s on the security variable. Table 3 contains the means and standard deviations. The summary table (Table 2) also shows a significant difference ($p < .025$) between fifth- and sixth-grade *S*s. It is interesting to observe, however, that the fifth-grade *S*s appear to be more secure ($M = 9.16$) than the sixth-grade *S*s ($M = 10.90$). One might speculate as to whether or not the onset of pubescence in sixth-grade *S*s contributes to this unexpected finding or whether the anticipated change to Junior High School is a factor contributing to insecurity. The grade by causality interaction effect is nonsignificant. Independent t tests for fifth- and sixth-grade *S*s (Table 3) result in significant differences for both groups. The data in Table 3 further show that homogeneity of variance may be assumed. Thus, we feel justified in concluding that high causally oriented *S*s are more secure than low causally oriented *S*s as measured by the Kooker Security-Insecurity Scale.

The analysis of variance summary relating to the second hypothesis—that high causally oriented *S*s have less anxiety than low causally

Table 2: *Summary of analysis of variance of insecurity, anxiety and L scores* ($N = 259$)

Types of Scores and Source of Variance	*df*	Mean Squares	*F*	*p*
Insecurity Scores				
Grades (G)	1	187.76	5.25	< .025
Causality (C)	1	753.19	21.04	< .001
G × C Interaction	1	50.89	1.42	*NS*
Within cells (w)	250*	35.79		
Anxiety Scores				
Grades (G)	1	233.05	5.62	< .025
Causality (C)	1	862.22	20.80	< .001
G × C Interaction	1	.93	.02	*NS*
Within cells (w)	253**	41.45		
L Scores				
Grades (G)	1	.02	.004	*NS*
Causality (C)	1	132.20	26.73	< .001
G × C Interaction	1	6.57	1.33	*NS*
Within cells (w)	253**	4.95		

* *Five degrees of freedom deducted from* df *for error since five mean values were added in order to obtain proportionality.*

** *Two degrees of freedom deducted from* df *for error since two mean values were added in order to obtain proportionality.*

Table 3: *Comparison of the mean scores of the high and low causally oriented groups on the Kooker Security-Insecurity Rating Scale, the Manifest Anxiety Scale, and the L Scale*

	High Causal Group			Low Causal Group			
	N	*M*	*SD*	*N*	*M*	*SD*	*t*
Security Scores[a]							
5th-grade *Ss*	60	8.15	5.26	45	10.51	5.86	2.11*
6th-grade *Ss*	88	9.10	5.22	66	13.29	7.27	3.94**
Total	148	8.72	5.22	111	12.16	6.81	4.42**
Anxiety Scores[b]							
5th-grade *Ss*	60	11.63	5.65	45	15.47	6.23	3.21**
6th-grade *Ss*	88	9.81	5.94	66	13.39	7.67	3.13**
Total	148	10.55	5.86	111	14.23	7.13	4.42**
L Scores[c]							
5th-grade *Ss*	60	2.10	1.96	45	3.93	2.44	4.09**
6th-grade *Ss*	88	2.40	2.25	66	3.58	2.22	3.22**
Total	148	2.28	2.13	111	3.72	2.30	5.14**

[a] *Low Kooker scores imply greater security.*

[b] *High scores reflect high anxiety.*

[c] *High scores indicate many L responses.*

* *Significant at the .05 level.*

** *Significant at the .005 level.*

oriented *Ss*—is also reported in Table 2. There is a significant ($p < .001$ difference between the high causally and the low causally oriented groups on the anxiety variable. As Table 3 indicates the difference is in the predicted direction, high causally oriented *Ss* manifest less anxiety as measured by the CMAS than low causally oriented *Ss*. Again there is a significant difference ($p < .025$) between fifth- and sixth-grade *Ss*. However, the trend on the anxiety scale is in the opposite direction from that on the insecurity measure. Sixth graders are more insecure than fifth graders, while sixth graders are less anxious than fifth graders. This finding

Table 4: *Intercorrelations of the criterion variables and IQ for fifth-grade ($N = 137$) and sixth-grade subjects ($N = 232$)*

	Security	Anxiety	L Scores
IQ			
5th grade	−.30**	.09	−.40**
6th grade	−.28**	−.20**	−.13*
Security			
5th grade	—	.002	.10
6th grade	—	.13*	.07
Anxiety			
5th grade	—	—	.11
6th grade	—	—	.13*

* *Significant at the .05 level.*
** *Significant at the .01 level.*

throws some doubt on the fifth hypothesis which stated that anxiety and insecurity are positively related variables. The grade by causality interaction effect of CMAS scores is nonsignificant. Independent t tests for fifth- and sixth-grade *S*s demonstrate significant differences ($p < .005$) for both classes. Tests for homogeneity of variance show that the assumption of homogeneity of variance is justified. In conclusion, the evidence supports the hypothesis that high causally oriented *S*s are less anxious as measured by the CMAS than low causally oriented *S*s.

The third hypothesis predicted that there would be a relationship between causality and L responses on the L scale. The summary Table 2 demonstrates that there is a significant difference ($p < .001$) between the high causal and the low causal groups on the L variable. The score on the L scale is obtained by counting the L responses. Thus, a low L score might be interpreted as honesty, a high L score as faking, dishonesty, or as lacking self-insight. Inspection of the mean (Table 3) demonstrates that the differences are in the predicted direction. The mean L score for the high causal group is 2.28, for the low causal group it is 3.72. There is no significant difference between fifth and sixth grades and there is no significant interaction effect between grade level and causality. Independent t tests for fifth and sixth grades demonstrate significant differences between the high and the low causal group. Homogeneity of variance may be assumed since the variance ratio is nonsignificant. Thus, there is support for the hypothesis that high causally oriented *S*s give fewer L responses, as measured by the 11 item L scale, than low causally oriented *S*s.

The F ratios which show the interaction effect of grades by causality in Table 2 are nonsignificant and thus support the fourth hypothesis. This is equally true for all three variables: security, anxiety, and L score. This finding and the t values reported in Table 3 provide justification for accepting the hypothesis that in this study high causally and low causally oriented *S*s respond differently to such indices of mental health as insecurity, anxiety, and L responses, irrespective of whether the *S*s come from the fifth or the sixth grade.

Table 4 reports the intercorrelations of the three criterion variables, as well as the correlation between IQ and each of the variables separately for fifth- and sixth-grade *S*s.

There is only a small positive correlation of .13 between anxiety scores and insecurity scores for sixth-grade *S*s, which barely reaches the .05 level of significance. The magnitude of the correlation does not change when analyzed separately for boys ($r = .13$, N 120) and girls ($r = 16$, N 112), but with smaller N's the correlations are no longer significant. The correlation between anxiety scores and security scores for fifth-grade *S*s is, for all intent and purpose, zero. Thus, we do not have conclusive evidence that our measures of anxiety and security are positively related. The data might indicate that both tests measure different and unrelated aspects or traits. However, we might also consider an alternative explanation arising from the different nature of the two test instruments, namely, that the anxiety scale is a paper-and-pencil test, consequently acquiescence, faking, rationalizing, and a deliberate attempt to give the socially desirable response might operate in a systematic fashion, while the Kooker scale uses the rating of behavior by trained observers. Possible shortcomings of this scale such as rater bias and halo effect would be operative in a different fashion. Bruce (1958) using the same tests with sixth-grade *S*s obtained a correlation of .26 (N 184) significant at the .01 level, thus giving some indirect support for our hypothesis. Furthermore, his finding is basically in agreement with our data obtained for sixth-grade *S*s. All that can be said at this point is that there are sufficient indications to warrant further investigation of the relationship between anxiety and insecurity as stated in our fifth hypothesis.

The correlation between insecurity scores and L scores is nonsignificant for both grade levels. The correlation between anxiety scores and L scores is nonsignificant for fifth graders and barely reaches the .05 level of significance for sixth graders. If the sixth-grade data are broken

Table 5: *Comparison of the mean scores on the Kooker Security-Insecurity Scale, the Manifest Anxiety Scale and the L Scale for causally and low causally oriented subjects with matched IQ's* (N = 100)

	High Causal Subjects				Low Causal Subjects		
	N	*M*	*SD*	N^a	*M*	*SD*	t^b
Security Scores	50	8.60	5.51	49	12.35	7.85	3.05***
Anxiety Scores	50	11.62	6.19	49	14.06	8.04	1.79*
L Scores	50	2.48	2.17	49	3.59	2.38	2.55**

[a] *The security score for one child, the anxiety and L scores for another child were not available.*
[b] *Since individuals are matched on the basis of IQ the standard error of a difference was computed for correlated means.*
* *Significant at the .10 level.*
** *Significant at the .02 level.*
*** *Significant at the .005 level.*

down into sexes the correlations are for boys −.11 (N 120) and for girls −.21 (N 112), both nonsignificant. Thus, the correlations of the anxiety scores with the L scores are not in disagreement with those reported by Castaneda et al. (1956) who found for sixth-grade boys $r = -.10$ (N 65) and for girls $r = .22$ (N 49), both nonsignificant, even though he reports a positive correlation for girls, while ours is negative.

As was indicated previously, the relationship between IQ and the criterion variables is relatively low and not always consistent for fifth- and sixth-grade *S*s. The correlation of the Kooker Security-Insecurity Rating Scale and IQ is −.28 for sixth- and −.30 for fifth-grade *S*s, both significant at the .01 level. Thus, there is a tendency for the more intelligent child to be rated as more secure. However, a correlation of .30 accounts for only 9% of the variance.

The correlation between anxiety scores and IQ is −.20 (significant) for sixth graders and +.09 (nonsignificant) for fifth graders, thus not only small in actual size but operating in different directions for the different grade levels. McCandless and Castaneda (1956) report correlations between IQ and anxiety scores for sixth-grade boys −.16 (N 55), nonsignificant, and for sixth-grade girls −.43 (N 45), significant.

The L scores show a moderately high significant correlation ($r = -.40$, N137) with IQ for fifth-grade *S*s, but only a barely significant correlation ($r = -.13$, N232) for sixth-grade *S*s. Again we feel justified in concluding that only a small amount of the variance is attributable to differences in IQ's. For a correlation of .40 only 16% of the variance can be accounted for in terms of IQ. However, there is a small tendency for the less intelligent children to give more L responses than the more intelligent *S*s.

In order to further eliminate the influence of IQ on the data, *S*s from the high causal group were matched with *S*s from the low causal group on the IQ variable. Sixth-grade *S*s were paired with sixth graders and fifth graders with fifth-grade *S*s. The IQ score difference between each pair was never greater than plus minus one IQ point. The data in Table 5 are based on 31 matched pairs of sixth graders and 19 matched pairs of fifth graders. The t test between the high causal and low causal *S*s matched on IQ is computed by way of the standard error of a difference between correlated means. The data indicate that IQ has no effect on the security data. The difference between the means is significant at the .005 level. The group means for the high and low causal *S*s are almost identical for the total 259 *S*s (Table 3) with the group means for the 100 *S*s who are used in the pairing (Table 5). Intelligence scores do not influence the results obtained in this study with the Kooker Security-Insecurity Scale.

For the anxiety data—even though the correlations between IQ and anxiety scores are lower and less consistent in magnitude and in direction for fifth- and sixth-grade *S*s than for the security scores (Table 4)—the t ratio between matched groups is not significant, but approaches the .05 level of significance. However, the correlation between IQ and anxiety scores for the 99 paired *S*s only is −.13, not significant. Thus, even though IQ scores appear to have some influence on the obtained differences on the anxiety data, this influence appears to be negligible.

The difference between the high causal and low causal *S*s matched on IQ for the L scores is significant at the .02 level. The L score means for the matched groups (Table 5) are about

Table 6: *A comparison of fifth- and sixth-grade subjects from causal and regular classrooms with respect to the selection variable*

	Causal	Regular	Total	Chi Square	p
Fifth Grade					
High Group	44 (29.2)[a]	15 (29.8)	59		
Low Group	8 (22.8)	38 (23.2)	46		
Total	52	53	105	31.65[c]	< .001
Sixth Grade[b]					
High Group	61 (46.5)[a]	32 (46.5)	93		
Low Group	23 (37.5)	52 (37.5)	75		
Total	84	84	168	18.88[c]	< .001

[a] *Figures in parentheses are expected frequencies.*

[b] *Reproduced with permission from Child Development Publications from a previous publication of the author (Muuss, 1960).*

[c] *Chi Square with Yates correction for continuity.*

the same as the means for the total groups (Table 3).

We thus can conclude that the results obtained in this study cannot be explained on the basis of differences in intelligence which existed between the original high causal and low causal groups.

So far the discussion has dealt only with the relationship between the degree of causality and indices of mental health, such as insecurity, anxiety, and L responses. No attention has been directed to the effects of a causal learning program in producing causally oriented *S*s. In line with the sixth hypothesis one might ask: To what extent did the experimental classes contribute subjects to the high and the low causally oriented groups? As noted in Table 6 chi square analysis indicates for both fifth and sixth grades a preponderance of cases from the experimental (causal) classes in the High Causal Cell and a preponderance of cases from the control (regular) classes in the Low Regular Cell, far beyond the number of expected frequencies.

Bruce (1958) demonstrated that sixth-grade children who had participated for two consecutive years in an experimental learning program designed to develop a causal understanding were more secure and less anxious than control *S*s who had not participated in such a program. He used the same tests which were used in this study but did not report the L scores. It is interesting to observe that he obtained significant differences only between the two-year group and the control group, but not between the one-year group and the control group. In this study a comparison was made between high causally and low causally oriented *S*s—a method more sensitive to differences than the comparison of experimental and control groups, but yielding basically the same results.

SUMMARY

Two tests, the Social Causal Test and the Physical Causal Test, served as selection criteria for high causally ($N = 148$) and low causally ($N = 111$) oriented *S*s. A comparison was made between the high causal and the low causal fifth- and sixth-grade groups in respect to measures of security, anxiety, and L responses. Data were analyzed by a factorial analysis of variance design followed by t tests. The findings of the study may be summarized as follows:

1. Fifth- and sixth-grade *S*s who are high causally oriented respond to measures of security, anxiety, and L responses in the direction that might be considered as indicative of mental health. The high causally oriented subjects show more security, less anxiety, and give fewer L responses than low causally oriented *S*s.
2. The obtained differences are equally obvious for the fifth- ($N = 105$) as well as for the sixth-grade *S*s ($N = 154$).
3. The differences between grade levels are inconsistent: fifth graders are more secure, sixth graders are less anxious. There is no between grade level difference on the L scale.
4. For sixth-grade *S*s there is no correlation, for fifth-grade *S*s there is a very low but significant correlation between insecurity and anxiety.
5. Even though there are small but significant relationships between intelligence and the criterion variables, the obtained difference between the high causal and low causal groups on the

criterion variables cannot be explained on the basis of differences in intelligence.

6. The experimental classes designed to develop a causal understanding of the dynamics of human behavior contribute significantly more high causally oriented *S*s, while the regular control classes contribute more low causally oriented *S*s to this study.

REFERENCES

BRUCE, P. Relationship of self-acceptance to other variables with sixth-grade children oriented in self-understanding. *J. educ. Psychol.*, 1958, *49*, 229–238.

CASTANEDA, A., McCANDLESS, B. R., & PALERMO, D. S. The children's form of the Manifest Anxiety Scale. *Child Develpm.*, 1956, *27*, 317–326.

CLARK, C. A. A study of the adolescent's conception of the nature of scientific knowledge. Unpublished doctoral dissertation, State University of Iowa, 1953.

KOOKER, E. An investigation of security, insecurity, achievement, and boredom in elementary school children. Unpublished doctoral dissertation, State University of Iowa, 1951.

LYLE, W. H., JR., & LEVITT, E. E. Punitiveness, authoritarianism, and parental discipline of grade school children. *J. abnorm. soc. Psychol.*, 1955, *51*, 42–46.

McCANDLESS, B. R., & CASTANEDA, A. Anxiety in children, school achievement, and intelligence. *Child Develpm.*, 1956, *27*, 379–382.

MUUSS, R. E. A comparison of "high causally" and "low causally" oriented sixth-grade children in respect to a perceptual intolerance of ambiguity test. *Child Develpm.*, 1960, *31*, 521–536.

OJEMANN, R. H., LEVITT, E. E., LYLE, W. H., & WHITESIDE, MAXINE F. The effects of a "causal" teacher training program and certain curricular changes on grade school children. *J. exp. Educ.*, 1955, *24*, 95–114.

STILES, FRANCES S. Developing an understanding of human behavior at the elementary level. *J. educ. Res.*, 1950, *43*, 516–524.

THE IMPACT OF THE CULTURE

45

A Cross-National Study of Children: A Study in Creativity and Mental Health

Harold H. Anderson and Gladys L. Anderson

The Andersons are concerned with the effects of what might be called a "national personality type" on children's responses to open-ended stories. Their presentation follows a highly logical pattern: they state their assumptions about the effects of democratic or dictatorial cultures on the child's ability to develop creative social relationships, they give the hypotheses deduced from their assumptions, and they conclude with the design and results. This report is one in a series of studies using these incomplete stories.

Aim. The purpose of this research program has been to develop and test an instrument that would be sensitive to cross-national similarities and differences and to examine children's responses in the light of certain hypotheses.

Materials. The *Anderson Incomplete Stories,* a

A Mimeographed Paper presented at the Sixth International Congress on Mental Health, Technical Session, 1961 (Paris, France), adapted from other research reports in preparation.

projective technique, were devised to elicit fantasy material from the child. Series A contains six short incomplete stories: each story presents a social conflict. Two stories present a conflict between a child and a parent; two, a conflict between a child and a teacher; and two, a conflict between one child and another child. In each story the child has made a mistake. Series B presents five such stories, in three of which an adult (father, mother, or teacher) has made a mistake. The stories were designed to represent experiences common to children in the Western cultures.

Administration. The administration of each series as a group test requires one school period. The teachers do not participate in the administration, and both children and teachers are told that the teacher and principal will not see the stories which the children write. The children are told not to sign their names. The children are told further that there are no right answers and no wrong answers, and that they will not be marked for these papers. They are asked to complete the stories as they honestly think the stories will end.

Sampling. Samples of school rooms of children about equally divided between boys and girls, contain children from high, middle, and low socioeconomic levels as determined by school authorities.

Subjects. The original purpose was to test the instrument with children in Germany and in the United States. A first sampling was gathered in Karlsruhe in 1952; a second sampling in Knoxville, Tennessee. Following a pilot study in Braunschweig, Germany, additional samples were obtained in Hamburg and Munich in 1954. In the United States additional samples have been obtained from Benton Harbor, Michigan, in the north; Cedar Rapids, Iowa, in the middle west; and San Juan, Puerto Rico, where the language is Spanish.

The inter-cultural scope was broadened by samplings from Birmingham, England; Drammen, Norway; Stockholm, Sweden; Helsinki, Finland; Mexico City and three Indian villages of rural Mexico: Tecómitl, Tepoztlán, and Chiconcuac; and lastly, in 1959, in Rio de Janeiro, Brazil.

The *Anderson Incomplete Stories,* in seven languages, have now been administered to over 10,000 children in the fourth and seventh grades in these eight countries. Data from seventh grade children are used in this report. . . .

Translation. The children's story completions have been translated into English.

Content analysis and coding. Separate coding manuals have been devised for content analysis of the children's story completions. Independent coders have shown percentages of agreement of over 90. Three of the stories are included in this report. The categories which are discussed in this paper were selected because they were regarded as important from the standpoint of values in human interacting, and therefore related to concepts of Creativity and to Mental Health.

Creativity and cross-national research. This cross-national study, begun in 1952, has since 1957 been a part of a research program in Creativity supported by a research grant. We are interested in Creativity not so much from the standpoint of painting and sculpture, as from the point of view of social creativity in human relations (Anderson & Anderson, 1954, 1961). There are different ways of regarding Creativity. From our perspective of personality development and productivity, Creativity is first to live truthfully. It is to live and to communicate truthfully as one himself sees the truth. To live merely according to the truth as anyone but himself sees the truth is not Creativity, but conformity or defense. We regard these statements as axiomatic: that they hold for Plato and Picasso, for Michelangelo and for Einstein. From these axioms certain corollaries follow. To the extent that behavior is characterized by truth and honesty it is undistorted by the needs for defense. It can thus be said to have a high validity by an internal criterion: honesty means that within the limits of one's own perceptions, one is what he purports to be. Creativity is also to live harmoniously, without distortions of threat or defense and with freedom to communicate validly.

The environment can become a facilitator of Creativity mainly in two ways, both essential: (a) by *accepting* the person as he is; and (b) by *stimulating* him through the confronting and free interplay of differences. This is our definition of democratic relating. That is, the humans in the environment facilitate Creativity in the person to the extent that they permit him to live harmoniously in an open environment where new ideas can emerge and can be honestly communicated, and to the extent that they stimulate him in non-threatening ways.

Persons in the environment restrict Creativity in the individual to the extent that they do not *accept* him and furthermore to the extent that they do not *stimulate* him through the free interplay of differences. In a restricting environment, instead of an Open system of interrelating there is a Closed system of environmental demands, threats, guilt and tabus, and other kinds of personal encroachments and usurpations.

The corollaries can be restated to relate more specifically to our cross-national study. By definition a closed restricting environment (Closed System) is to be found in greater extent in cultures traditionally known as dominating or authoritarian. By definition, also, the more Open System which fosters Creativity will be found in those cultures regarded as less authoritarian or, as we might say, more democratic.

Hypotheses. The hypotheses grew out of previous research on democratic and dictatorial teacher-child relations in American school rooms. It was assumed that the same differences in child behavior found in American school rooms would be found to exist in more extensive samplings of democratic and dictatorial cultures. The principal hypotheses of the cross-national research may be stated as follows:

(1) Children brought up in a more authoritarian and dominating culture are different in their inter-personal relations from children in less dominating (more democratic) cultures.

(2) Children in a more authoritarian culture will show *significantly higher* frequencies of expressions of defense mechanisms: i.e., responses of anxiety, lying, cheating, deception, daydreaming, escape from reality, ambiguous unstructured relating, conformity, guilt, submission, and also of aggressive, intolerant, and punitive behavior. In our scale of values these expressions represent one of two things, or both: a lower level of mental health in the person, or a less propitious environment for mental health, or both.

(3) Children in an authoritarian culture will also show *significantly fewer* responses of honesty, responsibility, sense of fair play, social problem solving, communication, spontaneity, cooperation, and harmonious, integrative behavior. These qualities of behavior are considered basic to a concept of Creativity and of social invention in human relations. They are also basic to high levels of mental health. Expressions of such behavior are expected in greater frequencies in democratic or less authoritarian cultures.

(4) It was assumed that samplings from Germany would be higher in reflecting an authoritarian culture and that samplings from England and the United States would be lower. After our first work in Mexico we learned that Mexican psychiatrists and psychologists frankly and openly regard their culture as highly authoritarian.

Values. We assumed that the children's story completions would reflect the choices that children and adults make in the processes of human interacting. In this sense the choices represent values, those qualities in human relating which the person regards as necessary, important, or traditional.

In previous analyses of our data from three German locations (Anderson & Anderson, 1960; Anderson, *et al.*, 1959; Erickson, 1961; Levitov, 1960), Hamburg, Karlsruhe, and Munich, we have found rather consistently that when the Hamburg children differed from children in Karlsruhe or Munich, the Hamburg children tended to write stories more like those from Birmingham, England, or Benton Harbor, Michigan. That is, Hamburg data are interpreted as reflecting more democratic values than are found in Karlsruhe or Munich.

FINDINGS

The Teacher's Money Story

The Teacher's Money Story reads as follows:

> The teacher suddenly discovers that fifty cents has disappeared from her desk. She looks up and sees that all the class are working on their arithmetic. She wonders what happened to the money and what she should do. What does the teacher do? *Finish* this story with a few sentences. Tell what happened to the money and exactly how the teacher feels and what she does.

Data are reported here from 4,636 seventh grade children in the eleven locations shown in Table 1. What, in fact, happened to the money in these eleven locations? Table 1 gives the percentages of children in each location who wrote that the money was lost, that is, the money was found under papers on the desk, it fell on the floor, or the teacher recalled some knowledge of spending it.

Specific hypotheses were that in the more democratic cultures there would be higher percentages of children reporting the money lost; in the more authoritarian locations the percentages would be lower. In Birmingham, England, 59.6% of the children wrote stories in which the money was lost. In the upper half of the distribution of Table 1 are the two Scandinavian locations of Helsinki and Stockholm, and the two continental United States samplings from Knoxville, Tennessee, and Benton Harbor, Michigan.

Table 1: *Money was lost*

N*	Location	%
398	Birmingham, England (Bhm)	59.6
208	Helsinki, Finland (Hel)	41.8
222	Knoxville, Tennessee (Knx)	38.3
166	Stockholm, Sweden (Stk)	33.2
282	Benton Harbor, Michigan (BnH)	32.2
349	San Juan, Puerto Rico (SnJ)	31.5
501	Hamburg, Germany (Ham)	30.9
1067	Karlsruhe, Germany (Kar)	24.3
332	Munich, Germany (Mun)	22.9
891	Mexico City, Mexico (Mex)	21.2
220	Drammen, Norway (Drm)	20.9
4636		

Table 2: *Money was stolen*

Loc.	%
Drm	54.6
Kar	44.8
Stk	44.0
Hel	40.4
Ham	40.1
BnH	40.1
Knx	36.0
Mun	34.9
Bhm	33.5
Mex	29.6
SnJ	18.6

* *All children were in the seventh grade.*

San Juan, Puerto Rico, was at the median, followed closely by Hamburg.

In the lower half of the distribution, hypothesized as authoritarian, were the three German samplings and Mexico City. Drammen, Norway, was lowest.

There is time here to make only this comment regarding the Norway sample. Leif Terdal and Gladys Anderson (1961) have presented analyses of data on this and several other stories, which show that the Drammen children have presented highly authoritarian responses when writing about teacher-child situations, but that when writing about parent-child and neighborhood situations, they were by contrast high in the rank orders of democratic responses.

Terdal and Anderson showed in one other location, San Juan, Puerto Rico, an inconsistency between the parent-child and teacher-child situations. In just the reverse of the responses from Drammen, San Juan children reflected a democratic human relationship in the school situation, but a contrasting high authoritarianism when writing about the home. In this respect San Juan children, who are both United States and Latin-American, differed from the continental United States children and from the children in Mexico City.

In Table 2 Drammen, with 54.6% led by ten per cent the children in all locations who wrote stories in which the money was in fact stolen. Note that the two Latin-American locations are at the bottom of the distribution, with San Juan 11 percentage points below Mexico City. To understand this position, we must look at Table 3.

The percentages of stories for each location in Tables 1 and 2 have been combined in Table 3 to compare and contrast the stories which were definite and explicit as to what in fact happened to the money with stories in which the child was vague, indecisive, uncommunicative.

It has been hypothesized that the freedom to think for oneself, the ability to communicate honestly and validly are prerequisites for creativity and for mental health.

The two Latin-American samples of Mexico City and San Juan, Puerto Rico, had virtually identical percentages in Table 3. Forty-nine per cent of the children in these two locations did not commit themselves; they wrote stories that were vague as to what happened to the money. There are many precautions necessary in interpreting the results of any projective technique, or in interpreting single or fragmented responses. It is our inclination, however, to interpret this vagueness as a culturally imposed fear response, as a reflection of the cultural traumatizing of children.

Next in rank order of vagueness of stories were the three German samples, with Munich children separated from Karlsruhe and Hamburg by 12 percentage points. It should be noted here that 25% of continental United States children have been thus culturally traumatized.

Conspicuous, and significantly different from the children in all other locations were the children in Birmingham, where only 7% were vague;

Table 3: *Authors' definite and vague statements: fact of definite loss or vague theft*

Loc.	%	%
Bhm	93.0	7.0
Hel	82.2	17.8
Stk	77.2	22.8
Drm	75.5	24.5
Knx	74.3	25.7
BnH	72.3	27.7
Ham	71.0	29.0
Kar	69.2	30.8
Mun	57.8	42.2
Mex	50.8	49.2
SnJ	50.1	49.9

Table 4: *Teacher's initial assumption*

A Assumes loss		B Assumes theft	
Loc.	%	Loc.	%
SnJ	35.5	Hel	74.5
Knx	31.6	Mun	73.7
Stk	24.7	Kar	73.4
BnH	22.0	Drm	71.8
Drm	20.9	BnH	71.3
Ham	20.2	Mex	71.1
Bhm	19.6	Stk	69.9
Mun	19.3	Ham	67.9
Mex	18.9	SnJ	57.3
Hel	15.4	Knx	56.3
Kar	14.8	Bhm	54.2

Table 5: *Teacher asks open question*

Loc.	%
Bhm	32.4
SnJ	28.7
Knx	28.4
Hel	24.5
BnH	22.0
Drm	20.5
Ham	19.0
Stk	18.7
Mun	17.2
Kar	17.0
Mex	12.1

93% were frank and explicit in making up their minds and in communicating definitely what happened to the money.

Several cultural consistencies can be noted in Table 3: the two Spanish-language locations have identical percentages; next are clustered the three German-speaking locations of Munich, Karlsruhe, and Hamburg; the two continental United States locations are adjacent and almost identical; the three Scandinavian locations cluster together on the categories of definiteness and vagueness.

What did the children reveal as the teacher's initial assumption concerning what happened to the money? Table 4 gives percentages of children who wrote stories in which the teacher initially assumed that the money was lost, or that the money was stolen. It was hypothesized that in the more democratic cultures there would be higher percentages of stories in which the teacher assumed that the money was lost. Though both the range and the percentages in Table 4 are relatively low, this hypothesis is generally supported.

It was also hypothesized that in the more authoritarian cultures the percentages of stories in which the teacher assumed theft would be higher. This hypothesis, too, is generally supported. Of general mental hygiene interest is the fact that such large numbers, ranging from 54.2% to 74.5%, wrote stories in which the teacher assumed that the money had been stolen.

There were striking discrepancies between the teachers' assumptions and the facts of the story as to what happened to the money, which would justify further analysis. For example, 59.6% of the Birmingham stories reported that the money was lost, whereas, only 19.6% of the children wrote that the teacher assumed loss.

The percentages of stories in which the teacher asked an open question are given in Table 5. The range and the percentages were generally low but consistent with the hypotheses concerning democratic and authoritarian relating.

The Football Story

The Football Story reads as follows:

> George and Tom are playing with a football. They know that they should not play football on the narrow space in front of the house. George gives the football a strong kick and the ball hits a window and makes a big crack in it. Tom thought that someone had appeared behind the window. No one could have seen who had kicked the ball against the window. *Finish* this story with a few sentences. Tell how the two boys felt about it and what they did.

Completion of the *Football Story* by the child requires choices involving certain cultural expectations or values. What is the intentionality toward the owner of the window of the boy who kicked the football? Does the kicker intend to communicate with the owner or will he run, hide, escape, and exclude the owner? What also is the intentionality of the Companion? Are the

Table 6: *Kicker's intent toward owner*

	A To communicate		B Not to communicate	
N	Location	%	Location	%
220	Knx	71.8	Kar	63.6
167	CRT*	65.3	Tcl	61.6
285	BnH	63.8	Mex	56.8
223	Drm	60.1	Mun	56.4
209	Hel	55.0	Tpz	55.3
158	Stk	52.5	Ccc	54.8
435	Bhm	50.4	MuS	54.0
218	Rio*	44.9	Ham	49.7
531	SnJ	41.6	Stk	43.1
278	MuS*	41.5	Bhm	42.1
487	Ham	40.6	Hel	40.2
328	Mun*	35.1	SnJ	38.8
1153	Mex	29.1	Rio	34.4
821	Kar	26.4	Drm	32.2
62	Ccc*	8.1	CRT	27.6
94	Tpz*	7.4	BnH	27.4
99	Tcl*	7.0	Knx	17.4
5768				

** Locations in addition to those in Table 1, all in seventh grade except CRT:*
CRT: Cedar Rapids, Iowa, total sample, sixth grade.
MuS: Munich, Germany, 1960 from same schools as Munich 1954
Mun: Munich, Germany, 1954
Tpz: Tepoztlán, Mexico (rural)
Rio: Rio de Janeiro, Brazil
Tcl: Tecómitl, Mexico (rural)
Ccc: Chiconcuac, Mexico (rural)

Table 7: *Restitution by boys*

A Voluntary		B Involuntary		C None	
Loc.	%	Loc.	%	Loc.	%
Knx	49.5	Bhm	23.2	Ccc	90.4
BnH	48.7	Mun	17.4	Tpz	87.3
CRT	47.4	MuS	16.9	Tel	85.9
Rio	29.8	Ham	16.2	SnJ	76.3
Drm	29.2	Drm	14.3	Kar	70.3
Hel	26.8	Kar	13.1	Mex	66.5
MuS	24.8	Hel	12.0	Rio	64.6
Bhm	24.6	Mex	11.7	Stk	61.3
Ham	23.4	Stk	11.4	Hel	60.3
Stk	23.4	CRT	10.2	Mun	59.8
Mex	21.4	Knx	9.5	Ham	57.5
Mun	20.1	BnH	9.1	MuS	55.7
SnJ	17.9	Tpz	7.4	Drm	55.6
Kar	14.9	Tcl	7.1	Bhm	51.5
Ccc	6.4	SnJ	5.3	CRT	42.0
Tpz	5.3	Rio	4.6	BnH	41.1
Tcl	4.0	Ccc	1.6	Knx	38.6

boys identified, and, if so, voluntarily or involuntarily? Or, not at all? Was anything said in the story about apology or restitution? Two categories are presented here: *Kicker's intent toward the owner*, and *Restitution by the boys*.

Table 6 shows for 5,768 children in seventeen locations percentages of stories in which the kicker intended to communicate to the owner of the window the knowledge of the damage. The category represented in Table 6A thus reflects a high cultural value. It represents an intent to communicate validly and honestly, to respect the owner, and, to this extent, to live harmoniously, all postulated for Creativity. Hypotheses were that in democratic locations there would be high percentages of communication; in authoritarian cultures the percentages would be lower. Cultural groups are seen to cluster again in Table 6A. The three continental United States samplings had the highest percentages. Next in rank order were the three Scandinavian locations. Conspicuously at the bottom were the three rural Mexican villages, while immediately above were Mexico City and the four German samplings.

In the analyses of stories from Rio de Janeiro, percentages tended to fall near the midpoint of most distributions, neither mainly democratic nor mainly authoritarian. Though Rio de Janeiro per cents are adjacent to those from San Juan in Tables 6A, 6B, and 7B, the children in Rio de Janeiro were much higher than San Juan children in percentages of stories showing the responsible behavior of *Voluntary restitution by the boys*, as given in Table 7A.

Table 6B gives per cents of stories in which the child wrote that the kicker intended not to communicate with the owner. In such stories the boys usually ran or tried to hide. The eight locations at the top of this rank order included the four Mexican and the four German locations. At the bottom of the distribution were the three continental United States locations, approached closely by the children from Drammen and Rio de Janeiro.

If two boys break a neighbor's window, do they make, or plan to make, restitution? Table 7A gives the percentages of stories in which the child wrote that restitution by the boys was voluntarily made, planned, or intended. Conspicuously separated as a cluster from all other locations were the three continental United States samplings at the top of the rank order. In sharp contrast at the bottom were the three Mexican villages. The remaining locations were not markedly differentiated.

In the stories reported in Table 7B the restitution was involuntary, that is, the boys were made to pay for the window. The percentages were relatively low. No explanation is offered here for the top ranking of Birmingham's 23.2%. Even with the small differences between loca-

Table 8: *Michael tells truth-lie*

Tells truth			Evades truth; tells lie		No information
N	Loc.	%	Loc.	%	%
288	BnH	61.0	CRT	15.9	28.8
225	Knx	60.0	BnH	25.7	13.2
398	Bhm	59.1	Knx	27.4	12.4
209	Hel	56.8	Bhm	32.4	8.5
225	Drm	56.0	Hel	36.3	6.7
170	CRT	55.3	SnJ	38.6	13.6
217	Rio	48.5	Drm	38.7	5.3
146	Stk	47.9	Rio	39.1	12.4
506	Ham	47.9	Tpz	42.2	34.4
413	SnJ	47.6	Stk	44.4	7.5
287	MuS	40.9	Ham	46.1	6.1
988	Mex	39.1	Mex	48.6	12.2
1098	Kar	37.8	MuS	50.9	8.0
335	Mun	36.5	Tcl	56.3	19.2
110	Tcl	24.5	Kar	56.9	5.2
99	Tpz	24.4	Mun	57.9	5.7
63	Ccc	19.1	Ccc	66.6	14.3
5777					

tions it can be noted that the three continental United States samplings had adjacent and almost identical percentages.

The *Football Story* presents a problem of negligence, well defined for adults in United States courts. The boys knew they should not play on this narrow space. When do children learn the basic principles of negligence, of harmony, of responsibility? Or, in different cultures, how many children have learned these principles at the age of thirteen when they are in the seventh grade?

Table 7C includes stories in which there was no restitution by the boys. The percentages were large, ranging from 38.6 for Knoxville, to 90.4 for Chiconcuac. If these story completions relate to human behavior, civilization has yet a long way to go. From the standpoint of values, the percentages in Table 7C favor the United States samplings, which had the lowest percentages. But in these United States samplings there were still 38%, 41%, and 42% of the United States children who wrote stories in which restitution was not made by the boys.

The Lost Meat Story

The Lost Meat Story reads as follows:

Michael's mother sends him to the store to get one pound of wieners. On the way home he puts the package of meat on the edge of the sidewalk and plays for a short time with his friends. A shepherd dog darts forward quickly, claws half the wieners out of the package and rushes away with them. Michael wraps up the remaining wieners and takes them home. What does Michael say to his mother? What does his mother do? How does Michael then feel about it? Think about these questions, then *finish* this story quickly with a few sentences.

In this story the loss of the meat constitutes an injury to the family associated with two transgressions which Michael made while on an errand for his mother: he laid the package down; and he played with friends. Can a child make a mistake? If he makes a mistake can he live honestly and truthfully with his mother? Or, in these several locations, how many children wrote stories of valid and of invalid communication, of evasion, anxiety, doubt, conflict, and punishment?

Table 8 includes stories which have been listed in mutually exclusive clusters of *Michael tells truth, Michael evades truth or tells lie,* and the *Story gives no information* about whether he told his mother the truth or a lie.

Table 8 is based on the story completions of 5,777 children from seventeen locations. The range of percentages of stories in which Michael told the truth was from 19.1% to 61.0%. Benton Harbor, Knoxville and Birmingham showed vir-

Table 9: *Mother actively participating in a non-hostile manner*

A Mother's non-hostile confronting		B Story does not apply
Loc.	%	%
Drm	27.5	72.4
Knx	23.6	76.4
Hel	23.4	76.6
Stk	21.2	78.8
CRT	20.0	80.0
BnH	19.1	80.8
Rio	15.2	84.8
SnJ	13.1	86.9
Mex	12.5	87.5
Bhm	12.1	87.9
Kar	8.7	91.2
Ham	8.0	92.1
MuS	8.0	92.0
Tcl	4.5	95.5
Mun	3.8	96.2
Ccc	3.2	96.8
Tpz	2.2	97.8

tually identical percentages of 61, 60, and 59. The remaining percentages generally confirmed the hypotheses concerning the higher expected frequencies of valid communication or of telling the truth in the allegedly democratic locations, and the lower frequencies of valid communication in the allegedly authoritarian locations. Consistent with other findings for the German locations, Hamburg had the highest per cent.

Among the percentages of stories in which *Michael evaded the truth or told his mother a lie*, the four lowest percentages were obtained in Cedar Rapids, Benton Harbor, Knoxville, and Birmingham, ranging from 15.9% to 32.4%. The percentages for the German locations ranged from Hamburg, 46.1% to Munich, 57.9%. Again, interspersed within the range for the German locations were Mexico City, 48.6% and the Mexican village of Tecómitl, 56.3%. Rio de Janeiro was at the median.

No interpretation is made here of the percentages of stories in which there was *No information* as to Truth-Lie.

The category called *Mother's non-hostile confronting*, shown in Table 9, includes stories in which the mother was actively participating, communicating, interacting with the son in a non-hostile manner. These non-threatening qualities of relating are postulated as being high in social value and as facilitating communication by the child. The frequencies were low, ranging in percentages from 2.2% to 27.5%. The six top-ranking locations, led by the children from Drammen, who have tended to reflect a democratic influence at home, included the two other Scandinavian locations of Helsinki and Stockholm and the three locations from continental United States.

The seven locations lowest in percentages of stories containing the *Mother's non-hostile confronting* included the four German samplings and the three Mexican villages. Table 9 thus presents additions to the internal consistencies found in the children's story completions.

DISCUSSION

An assumption basic to all projective techniques is that one's fantasy responses are derived from past experiences and reflect directly or symbolically one's perceptions, experiences, and needs. A limitation of all projective techniques is the problem of establishing in precise fashion the relation between fantasy and reality. Do children and adults behave differently in different locations, or do they merely write different stories? Are human relations really different in Benton Harbor, Hamburg, and Munich? Do the children, for example, in Benton Harbor tend to complete the stories mainly in the way in which they think adults would expect? Are they more conformists? Or are the children in Munich, on the contrary, more honest and realistic, and do they complete the stories more nearly as children and adults everywhere actually behave?

Whatever the relation between fantasy and behavior, there were large and consistent differences in the fantasy productions of children in these different geographical samplings. Important to note is the fact that these differences, however large, were relative: they represented not all-or-none, but more-and-less as expressed in percentages. If authoritarianism was reflected in the German and Mexican stories it was present also, though only in lesser degree, in all the other locations.

Projective techniques, as psychological instruments or measuring devices, are known to be low in reliability and low in validity. On the basis of the categories from our coding manuals, interpreted broadly in terms of values, the *Anderson Incomplete Stories* have not elicited erratic, whimsical, or chance responses. In these children's story completions there are, on the contrary, internal consistencies noteworthy for any projective technique.

REFERENCES

ANDERSON, H. H., & ANDERSON, GLADYS L. Social development. Chapter 19. In L. Carmichael (Ed.), *Manual of child psychology*. 2nd ed. New York: Wiley, 1954. Pp. 1162–1215.

ANDERSON, H. H., & ANDERSON, GLADYS L. A imagem do professor na mente das criancas em cínco países. *Anais do VI Congresso Interamericano de Psicologia, Rio de Janeiro, 1960*. Rio de Janeiro: Fundafão Getulio Vargas, 1960. Pp. 43–52.

ANDERSON, H. H., & ANDERSON, GLADYS L. Image of the teacher by adolescent children in seven countries. *Amer. J. Orthopsychiatry*, 1961, *31*, 481–520.

ANDERSON, H. H., ANDERSON, GLADYS L., COHEN, I. H., & NUTT, FRANCES D. Image of the teacher by adolescent children in four countries: Germany, England, Mexico, United States. *J. soc. Psychol.*, 1959, *50*, 47–55.

ERICKSON, D. J. A comparison of children's social values in story completions by culture-groups. Unpublished master's thesis, Michigan State University, 1961.

LEVITOV, EDITH SYLVIA. "The broken bicycle": A comparative study of story completions by adolescents in United States and German cities. Unpublished master's thesis, George Washington University, 1960.

TERDAL, L., & ANDERSON, GLADYS L. Variability in children's values in eight countries as a function of situational differences. Paper presented at Midwestern Psychological Association, Chicago, May 1961.

46

Children's Learning from Television

Wilbur Schramm, Jack Lyle, and Edwin B. Parker

The impact of television on children's morals, attitudes, behavior, and even eyesight has been the subject of continuous debate and investigation for the last decade. This study is particularly valuable because Schramm and his colleagues were able to conduct a field experiment using two communities comparable except for the possession of television.

No one doubts that children learn from television. Nearly 100 studies now testify that they learn efficiently in the classroom from instructional television. There is no similar group of studies to prove that they learn in the home from the non-instructional television to which they devote about one sixth of their waking hours from the ages of 3 to 16, but the fact of this learning has been generally accepted. No parent who has watched his children reproduce language and mannerisms, mention names, and beg boxtops, all familiar to him only through television, doubts that his children are learning from television. Indeed, the question of what a child learns from television has been second in the public mind only to the question of what a child learns in school; although whereas the worry has been that the child might not be learning *enough* in school, it is rather that he might learn *too much*, of an undesirable kind of knowledge, from television.

It may be useful, therefore, to try to sketch in some of the dimensions of a child's learning from home television. What does the child tend to learn from television? Does he get an earlier start, thanks to television, in learning about his environment? Does he get an earlier start in learning the skills of communication? Does television really change anything in the amount or nature of his knowledge, from what he would have known without television?

These are not easy questions to answer. For example, the Himmelweit-Oppenheim-Vince team who made a sophisticated study of children's television behavior in England (1958) were forced to admit: "We were not able to find a sufficiently large number of information items (despite careful monitoring of children's and early evening television) which, without being trivial or abstruse, were not also known equally well to the controls (children without television) —a further indication of the essential similarity of the material put over by the different mass media. . . . One or two items of rather specialized information included in the test showed that the viewers do pick up information—but their number was too small to permit more detailed analysis" (p. 29).

We are very fortunate, therefore, as a by-product of a recent study of television behavior, involving 6,000 children in the United States and Canada (Schramm, Lyle, & Parker, 1961),

---

Reprinted by permission from Edward C. Uliassi (editor). *Studies in Public Communication*. Volume 3 (series). Chicago: Department of Sociology, University of Chicago, 1961.

to be able to sketch in, at least tentatively, some of the outlines of children's learning from television. These are the findings we are here reporting.

THE SOURCES OF THE DATA

In the course of the large study mentioned above, we were able to give knowledge tests (along with a great many other instruments and questions) to three fairly large samples of children: (a) 2,196 children sampled from the first six, and the 8th, 10th, and 12th grades in San Francisco; (b) 1,708 children representing the 1st, 6th, and 10th grades in five Rocky Mountain communities; and (c) 913 children, representing the 1st, 6th, and 10th grades in two Canadian communities, one of which did and one of which did not yet have television.

These two latter communities are of special interest to the problem we are discussing. Therefore, let us note about them that they are both cities of about 5,000 persons, in the same general geographical and cultural area, supported by about the same industries and businesses, containing about the same ethnical and religious groups, and served by the same types of school system. But one of them, which we shall call Teletown, is so situated that it receives not only all the Canadian network television, but also most of the United States service. The other, because of the accident of geography and the delayed extension of cables and microwave systems, has no television reception. Teletown has television receivers in more than 75 per cent of its homes, and every child in our sample had regular access to television programs. In the other community, which we shall call Controltown, there was only one television receiver connected, and it received signals, the owners reported, about three nights a year. In these two communities, therefore, we have an opportunity to make comparisons of children's informal learning with the culture held relatively constant, so to speak, except for the presence or absence of television.

THE NATURE OF CHILDREN'S INFORMAL LEARNING FROM TELEVISION

Let us be clear that we are dealing now, not with what children learn from the formal and patterned use of instructional television in the classroom, but rather from the informal use of television in the home. This latter use represents at least 95 per cent of all the television they see.

It is evident that most of a child's informal learning from television will be *incidental* learning. By this we mean learning that takes place when a viewer goes to television primarily for entertainment, and stores up certain items of information without seeking them. Practically all of a child's early use of television is in quest of entertainment. Out of 111 programs named by San Francisco children in the first six grades as programs they liked well enough to try to hear regularly, only four were programs in which the information content, rather than the entertainment content, appeared to be central to the purpose of the program. The act of going purposefully to the media for useful information is behavior that is learned chiefly in school, and more often sends the child to print than to the audio-visual media. The idea of going purposefully to television at home, to seek information, is a relatively uncommon behavior, learned late, and, as we have been able to document in our study, likely to be engaged in, to any considerable extent, only by a relatively small group of children. Therefore, most of the learning that takes place from home television must be incidental to the main purpose of viewing, and we can focus our attention on the presumed characteristics of a child's incidental learning.

The amount of learning, we can suppose, will depend on abilities, needs, and focus of attention. The easiest of these to define and predict upon is mental ability. Bright children, other things being equal, will learn more than less bright ones. So we can suppose that the brightest children, who are likely also to be "achievers," can be expected to make maximum use of *any* opportunity to learn from television.

However, there is also reason to think that television might represent an effective learning opportunity for children of below-average mental ability. This has been found empirically in some of the studies of instructional television, but it may also be predicted theoretically on the grounds that television obtains from the slower children an extent and degree of attention which they do not commonly give to learning experiences. It supplies motivation they otherwise might not have. This should be especially noticeable in the early years of their lives before they approach the ceiling of their ability to learn.

There are certain other reasons, also, for expecting relatively more learning from television early in a child's life.

For one thing, we should expect a child to be more likely to pay attention to, and store up, some fact or item of behavior, if it is *new* to him. Brodbeck (1955) reports a study of a "Hopalong Cassidy" film which produced considerable learning and behavior change on the part of younger children, but very little on the part of older children. After some investigation, they concluded that this was because the film was a new experience to the young children, but "old stuff" to the older ones. This familiarity effect is especially important in television because of the astonishing repetitiveness of many of the program types. As a child becomes more familiar with the medium, he learns to sort out the material. He sets up a pattern of expectation which keeps him from having to pay too much attention to the familiar items, and allows him to concentrate on the new ones—if they are not *too* unfamiliar.

On this basis, we should expect that the greatest amount of incidental learning from television would take place, let us say, in the ages from three to eight, when television has the least competition. Then the child's slate is relatively clean. Almost any experience is new to him, and therefore absorbing. Later, the competition for attention grows more severe. Every year the memory shelves grow a little more full. But in those years before the child learns to read well, when his horizon is still narrow and his curiosity boundless, when almost everything beyond his home and his little family circle is new—that is the time television has a unique opportunity to contribute to his information and his skills.

Another reason why television should be an especially effective early agent of incidental learning is the matter of its seeming *real*. A number of observers have noted that mass media content has greater impact on children if they believe it "really happened." In the early years of childhood, there is a shadowy border at best between the story world and the real world, and the events of the picture tube, the screen, or the bedtime story often seem terribly real. As children grow older they begin to develop . . . the habit of the adult of looking at an entertainment film or television program as art rather than life. Therefore, during the years when television is still seen as real, and thus transferable to life, we can expect a broader range of learning. But let us be careful not to suggest that learning does not occur, or reality is not perceived, even in the midst of play-acting. We know that many a girl can get suggestions for her love life or her spring wardrobe, even from a program which she knows is drama rather than real life.

Thus far we have been talking about the children's side of the learning. It is obvious, however, that the individuals and organizations who produce programs have something to do with incidental learning from television, because they control what goes into the programs. To be sure, all children do not perceive the same things in the same program, and no program is perceived by the audience precisely as the author or producer conceived it. But on the other hand, television's own focus of attention will obviously have something to do with what is learned from it. Television's focus is overwhelmingly on fantasy and formula programs: westerns, whodunits, situation comedies, and so forth. These programs are quite redundant and repetitive. We have already noted that children typically choose few of the reality programs on television (meaning the programs intended to illuminate public affairs, science, and the like). Therefore, it is highly probable that the average child will be getting a heavy diet of formula and fantasy. In the first years of his viewing, all this will be new to him and we can expect that his incidental learning from it will be at a maximum. But as he becomes more accustomed to the programs, we can expect him to find fewer and fewer new items to learn. Therefore, on this basis too, we can expect a child to learn relatively less from television as he grows older. And if this program diet is predominantly formula and fantasy, we can predict that his incidental learning will be chiefly from this kind of material. . . .

THE HYPOTHESES

On the basis of what we have been saying, we should expect three relationships to hold, at least in general terms:

(1) Television should give children a faster start at learning useful skills and knowledge about their environments than children typically were given before the age of television (because of the great vividness of the medium, and the amount of time it commands in years when children's memories are least crowded); but that this advantage would be lost after a few years (because of the repetitiveness of television, because children use television relatively little as a source of reality information and because older children without television obtain comparable information from other sources).

(2) In this first advantageous experience with television, the brightest and the slowest children in mental ability, will benefit rather more than the average children, in comparison to children without television (because the brightest children are "achievers" and will make maximum use of the new medium, and because the very attractiveness of television will provide motivation and command time which the slower children would otherwise not give to a comparable learning experience).

(3) That after the first years of accelerated learning, the effect of television on a child's knowledge will be to increase the levels of knowledge in areas which television, as used by children, chiefly emphasizes (the world of entertainment and fantasy), to decrease for heavy viewers of television the levels of knowledge in areas more easily learned about from other media (such as literature and public affairs), and to make no difference in levels of knowledge in areas which are learned chiefly from sources other than the mass media (for example, science, which is learned chiefly in school).

PROCEDURE

The best setting in which we can test the first part of (1) and (2) is with standardized and special vocabulary tests which we gave in Teletown and Controltown. We can hypothesize, in regard to these, that first grade children in Teletown will have larger vocabularies than first grade children in Controltown. We can also hypothesize that the differences will be larger in the case of the brightest and slowest children than in the case of the average children.

We can test the second part of (1) with information tests we gave the sixth and tenth grades in Teletown and Controltown on such subjects as science, names in the news, famous authors, famous entertainers, and so forth. Here we hypothesize that there will be no significant difference between Teletown and Controltown children, in corresponding grades, on these tests. In other words, the Controltown children will have caught up with the Teletown children.

Our best opportunity to test (3) is by comparing heavy and light viewers in all three samples (Teletown, San Francisco, Rocky Mountain towns). Here we have numbers that more easily permit a test of significance. Therefore, we hypothesize that in the sixth grade and beyond, in these three samples with mental ability held constant, heavy viewers will score above light viewers on tests that clearly relate to the fantasy content of television, less highly than light viewers on tests that relate clearly to reality material learned more easily from other media than from television, and neither higher nor lower on tests that relate to areas not primarily learned from the mass media or learned equally well from all the media.

TESTS OF THE HYPOTHESES

We gave a Stanford-Binet vocabulary test, standardized on very large North American samples, to first grade children in Teletown and Controltown, in the same week of the school year. [Tables 1 and 2 show] the means of these tests, along with appropriate tests of significance, and an analysis of variance the results of which justify the *t* tests of significance.

It appears, then, that first grade children of high intelligence and low intelligence in a television community do score significantly higher than their counterparts in a non-television community, on general vocabulary. The difference amounts to about one school grade. Between children of average intelligence in the two communities, however, the difference is negligible.

We gave these first grade children also a special vocabulary test made up of five words which they might be particularly likely to encounter on television—the words being *satellite, Prime Minister, pistol, war,* and *cancer.* [Tables 3 and 4 present] the results of that test in the two towns.

Again, the analysis of variance justifies continuing with the *t* tests. These show that there is a significant difference in favor of the Teletown children among the lowest intelligence group, but that the difference in the high group does not quite reach significance, and, again, that the difference between average children in the two towns is negligible.

To press the analysis further, we also compared the performance of heavy and light viewers in Teletown on the two vocabulary tests. Analysis of variance calculations on the general vocabulary test results show a television effect significant at the .05 level, but none of the intelligence groups taken separately attains a significant difference, although the average group is near the .05 level. In the special vocabulary test, however, significant differences are found, and these are recorded in [Tables 5 and 6.]

Again we find that the highest and lowest

Table 1: *Mean general vocabulary scores for first-grade children in Teletown and Controltown, by intelligence grouping*

Intelligence	Teletown		Controltown		S[a]	t
	N	Mean	N	Mean		
IQ above 115	29	9.10	21	8.19	1.78	1.78*
IQ 100–115	57	7.77	89	7.85	1.63	—
IQ below 100	33	7.52	27	6.33	1.48	3.11**

[a] *Unbiased estimate of standard deviation (within groups)*
* *Significant beyond the .10 level (.05 by one-tailed test)*
** *Significant beyond the .01 level*

Table 2: *Results of analysis of variance calculations for general vocabulary scores for first-grade children in Teletown and Controltown, by intelligence grouping*

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F
Intelligence	2	59.98[a]	29.99	11.36***
Television	1	22.05[a]	22.05	8.35**
Interaction	2	21.43[a]	10.71	4.06*
Individuals	250	659.31	2.64	—

[a] *Corrected for disproportionate subclass frequencies*
* *Significant beycnd the .05 level*
** *Significant beyond the .01 level*
*** *Significant beyond the .001 level*

Table 3: *Mean special vocabulary scores for first-grade children in Teletown and Controltown, by intelligence grouping*

Intelligence	Teletown		Controltown		S[a]	t
	N	Mean	N	Mean		
IQ above 115	29	3.10	21	2.71	1.01	1.00
IQ 100–115	57	2.68	89	2.52	1.11	—
IQ below 100	33	2.39	27	1.78	.93	2.52*

[a] *Unbiased estimate of standard deviation (within groups)*
* *Significant beyond the .05 level*

intelligence groups appear to benefit more from television than the average group, and that heavy viewers in both groups do significantly better than light viewers. This result gives us confidence that the differences we are finding are really related to viewing of television.

Therefore, we have considerable support for the hypotheses that first grade children in a town with television will have larger vocabularies than corresponding children in a town without television, and that the differences will be more pronounced in the highest and the lowest intelligence groups than in the middle group.

We are on less firm ground in testing the

Table 4: *Results of analysis of variance calculations for special vocabulary scores for first-grade children in Teletown and Controltown, by intelligence grouping*

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F
Intelligence	2	—	—	—
Television	1	5.85[a]	5.85	5.32*
Interaction	2	2.37[a]	1.19	—
Individuals	250	276.06	1.10	—

[a] *Corrected for disproportionate subclass frequencies*
* *Significant beyond the .05 level*

Table 5: *Mean special vocabulary scores for first-grade children in Teletown by intelligence grouping and by television viewing (Heavy viewing = 1 hr. 15 min. or more per average school day by parents' report)*

Intelligence	Heavy Viewers		Light Viewers		S[a]	*t*
	N	Mean	N	Mean		
IQ above 115	13	3.53	16	2.75	.95	2.23*
IQ 100–115	34	2.71	23	2.65	1.05	—
IQ below 100	19	2.68	14	2.00	.95	2.03*

[a] *Unbiased estimate of standard deviation (within groups)*
* *Significant beyond .05 level*

Table 6: *Results of analysis of variance calculations for special vocabulary scores for first-grade children in Teletown by intelligence grouping and by television viewing*

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F
Intelligence	2	—	—	—
Television	1	5.00[a]	5.00	5.00*
Interaction	2	3.16[a]	1.58	—
Individuals	113	112.61	1.00	—

[a] *Corrected for disproportionate subclass frequencies*
* *Significant beyond .05 level*

hypothesis that there will be no difference in knowledge levels in the sixth and tenth grades. Here we did not give vocabulary tests. Six tests and scales are available, however, using the sixth and tenth grade children in Teletown and Controltown. These are on science, ability to name bandleaders and popular singers, knowledge of literature, faraway place names, and statesmen and rulers. There was no over-all difference between the children of the two communities in these tests, in either of the grades, and very few significant differences in test performance even when the children are divided by intelligence and amount of viewing. This is precisely

Table 7: *Results with subject matter related to television content*

Test	Result	Significance
Ability to name singers	11 out of 13 comparisons favor heavy viewers; 1 is even; 1 favors light viewers	Significant beyond .01 level
Ability to name dance band leaders	8 out of 13 favor heavy viewers; 2 are even; 3 favor light viewers	Significant beyond .05 level

Table 8: *Results with subject matter learned from media other than television*

Test	Result	Significance
Ability to name writers	11 out of 13 comparisons favor light viewers; 2 favor heavy viewers	Significant beyond .02 level
Ability to name rulers and statesmen	10 out of 13 favor light viewers; 2 are even; 1 favors heavy viewers	Significant beyond .01 level

Table 9: *Results with content not related to mass media*

Test	Result	Significance
Ability to solve science problems	5 comparisons out of 10 favor light viewers; 2 are even; 3 favor heavy viewers	Not significant
Ability to recognize faraway place names	3 out of 9 favor light viewers; 4 are even; 2 favor heavy viewers	Not significant

the experience of Himmelweit, Oppenheim, and Vince (1958), previously referred to, who worked with children bracketing the ages of our sixth graders.

We can report some trends. When the children are divided into groups by school grade and three levels of intelligence, and each of these groups is compared with its corresponding group in the other town, on each of the six tests, we have a possibility of 18 comparisons in each grade, 36 in all. Over 60 per cent of these comparisons, when there is a difference, are in favor of Controltown. Except for one tie, *all* the comparisons in the sixth grade are in favor of Controltown. All the comparisons on the science test, when there is a difference, are in favor of Controltown. To be sure, very few of these differences are significant. However, by sign test, the performance of the Controltown children is significantly superior in the sixth grade, and there is no significant difference in the tenth grade. Thus, we have every reason to believe, at least until better data becomes available, that there is no advantage to sixth and tenth grade children in a television community in the kinds of knowledge here measured. If there is any difference at all, more learning of these kinds would seem to take place among comparable children who do *not* have television.

In order to test the third series of hypotheses, we used the San Francisco, Rocky Mountain, and Teletown samples. Even so, however, it seemed better not to combine the samples, because in San Francisco we had eighth, tenth, and twelfth grades, rather than sixth and tenth, as in the other samples; and furthermore, there were a few slight differences in the tests and scales between samples (for example, in Teletown the Canadian children were not asked to name U.S. Senators, as were the Rocky Mountain and San Francisco children). Therefore, we made comparisons within samples, dividing each grade by heavy and light viewing and by intelligence, and comparing heavy viewers with light viewers in the sixth grade Rocky Mountain sample, and so on. This required us to use sign test statistics rather than more usual comparisons of means.

The results are as predicted. In the case of subject matter closely related to television's central content, the results [are shown in Table 7.]

With respect to the subject matter which we judged could be learned more easily from media other than television, [the results appear in Table 8.]

Thus we have some evidence that television viewing may contribute to entertainment-related information, and not to information which relates to fine arts or public affairs and is thus more readily learned from print or some other medium.

We have fewer comparisons by which to test the hypothesis that there will be no difference in knowledge of subjects usually not learned from mass media (like science) or learned equally from several media (like faraway place names). Some of the tests were not given in all places, [but the results are given in Table 9.]

On this evidence, at least, we have no basis for believing that there is any real difference related to television in knowledge of this kind.

BRIEF DISCUSSION

The hypotheses hold up, and we have at least a tentative set of propositions for estimating the amount and kind of incidental learning that goes on as a by-product of children's two or three daily hours of television.

We feel it is necessary to say a word of caution about these results. For one thing, they are a fragmentary report on learning. They contain very little information about many areas of knowledge, or about learning to solve problems or think critically, or about the learning of taste and values, or about stimulation to undertake cultural or scientific or other desirable activities on one's own. To be sure, what we know or conjecture about most of these aspects of learning is not particularly favorable to the television experience, but it is important to know more about them before we close the books on the subject of learning from home television.

However, we have here some suggestive facts. Television *does* apparently send children off to a faster start in the skill they perhaps need most at an early age, although the advantage is chiefly to the brightest and the slowest children and not to the great group of average children. We can suppose, too, with some confidence that the same children who learn more vocabulary from television will also have learned more about their distant environments than comparable children without television. But the faster start is apparently frittered away in a repetitive diet of westerns, whodunits, and situation comedies; and it seems that heavy viewing in the sixth grade or beyond results in more knowledge (than light viewers have) of matters related to entertainment, but less knowledge of public affairs and fine arts.

These facts obviously have implications for public policy, although we have no intention of developing those implications here. However, we should like to suggest a few questions. If these conclusions are indeed broadly applicable, they would suggest that one of the great informative resources for young minds is being used at considerably less than what most observers would consider optimum. Is this necessary, considering the nature of the media and of its support, or can something be done so that the initial advantage of incidental learning from television can be broadened and extended in time? For example, there are many more reality opportunities on television than children make use of. Can parents by example, or schools by precept and interpretation, increase children's use of these reality experiences? To what extent are children's preferences for the richly supported fantasy and formula programs over the weakly supported reality programs inevitable considering the nature of children, and to what extent can they be traced to the support of these programs and the way they are made? With all the talent and resources of broadcasting, and, if necessary, with the resources and talents of foundations and educational institutions, is it possible to find out how to make reality programs skillfully enough that they will be truly competitive for children's interest with formula and fantasy? These are questions which obviously relate to the amount and kind of children's informal learning from television, and we suggest that they are questions with serious import for parents, schools, and broadcasters.

REFERENCES

BRODBECK, A. J. The mass media as a socializing agency. Paper for the American Psychological Association Symposium on Children and Mass Media, San Francisco, September 1955. Mimeo.

HIMMELWEIT, HILDE, OPPENHEIM, A. N., & VINCE, PAMELA. *Television and the child.* London: Oxford University Press (published for the Nuffield Foundation), 1958.

SCHRAMM, W., LYLE, J., & PARKER, E. B. *Television in the lives of our children.* Stanford: Stanford University Press, 1961.

PART V

ADOLESCENCE

PHYSICAL DEVELOPMENT

FAMILY INFLUENCES

THE PEER SOCIETY

PERSONALITY AND SELF

INTELLIGENCE AND CREATIVITY

CONCEPT DEVELOPMENT

PHYSICAL DEVELOPMENT

47

Self-Conceptions, Motivations, and Interpersonal Attitudes of Late- and Early-Maturing Boys

Paul Henry Mussen and Mary Cover Jones

As a phase of the work of the California Growth Study, Mussen and Jones investigated the relationship between the rate of one's physical growth and one's self-perceptions, and in this article they report the data on adolescent boys. Because this was a longitudinal study, it was possible to examine the relationship after physical maturity had been achieved. This article and the next reveal both some strengths and weaknesses of longitudinal studies.

While many intensive case studies show that personal and social adjustment during adolescence may be profoundly influenced by rate of physical maturation, there is a scarcity of systematic data on the relationship between the adolescent's physical status and his underlying motivations, self-conceptions and interpersonal attitudes. There is, however, a small body of evidence which demonstrates that greater physical maturity is associated with greater maturity of interest among girls and that early-maturing boys differ from their late-maturing peers in both overt behavior and reputational status. In one study (Jones & Bayley, 1950) in which a staff of trained observers assessed a large group of adolescents on a number of personality variables, boys who were consistently retarded in physical development were rated lower than those who were consistently accelerated, in physical attractiveness, grooming, and matter-of-factness; and higher in sociability, social initiative (often of a childish, attention-getting sort), and eagerness. Reputation Test (Tryon, 1939) data indicated that classmates regarded the late-maturing boys as more attention-getting, more restless, more bossy, less grown-up, and less good-looking than those who were physically accelerated.

On the basis of these findings, it may be inferred that adult and peer attitudes toward the adolescent, as well as their treatment and acceptance of him, are related to his physical status. This means that the sociopsychological environment to which late-maturers are subjected—and consequently the social learning situations they encounter—may be significantly different from that of their early-maturing peers. As a consequence, according to the ratings summarized above, they acquire different patterns of overt social behavior. It seems reasonable to hypothesize that groups differing in physical status will also differ in more covert aspects of behavior and personality.

Indirect evidence relevant to this hypothesis comes from an investigation of the long-term

---

From *Child Development*, 1957, *28*: 243–256. Reprinted with permission of the Society for Research in Child Development.

consequences of physical acceleration or retardation during adolescence. Jones (1957) found that group differences in physique had practically disappeared by the time her early- and late-maturing subjects reached their early thirties. Nevertheless, young adults who had been physically retarded adolescents differed from those who had been accelerated in several important psychological characteristics. In general, it appeared that the adult subjects could be described much as they had been during adolescence. Thus, those who had been early-maturers scored higher on the good impression, socialization, dominance, self-control (low score on impulsivity), and responsibility scales of the California Personality Inventory, while those who had been slow in maturing scored higher on the flexibility scale. On the Edwards Personal Preference Schedule, early-maturers scored significantly higher on the dominance scale, while the late-maturing were high in succorance. Jones concludes that the early-maturing "present a consistently favorable personality picture with regard to . . . important social variables" (1957). Moreover, there was some evidence that these men had attained more stable vocational adjustments than those who had been late in maturing. These group differences in later adjustment suggest that the sociopsychological atmosphere in which the adolescent lives may have profound immediate and enduring effects on his personality structure as well as on his overt behavior.

The present study was designed to investigate the relationship between maturational status and certain important, covert aspects of personality during late adolescence. Personality structure was assessed by means of the Thematic Apperception Test (TAT) which seems to be the most appropriate and sensitive instrument for this purpose. More specifically, on the basis of the literature reviewed above and other general works on the psychology of adolescence, we formulated and tested a series of propositions relating to differences between the physically retarded and the accelerated in self-conceptions, underlying motivations, and basic interpersonal attitudes. These variables were translated into TAT categories—needs (*n*), press (*p*), and descriptions (defined briefly in Table 1)—and the scores of early- and late-maturers in each of these categories were compared. The propositions and the rationale underlying them, together with the TAT variables involved, follow.

1. In view of their obvious physical retardation, relatively unfavorable reputations and disadvantageous competitive position in many activities, the late-maturing boys are more likely to have feelings of inadequacy. Hence, more boys in this group than in the early-maturing group are likely to have negative self-conceptions (TAT category: *negative characteristics*).

2. The adolescent in our culture generally desires some independence and adult status. This may be the source of a major problem for the late-maturer, however, since he is often regarded and treated as a small boy by adults and peers and is not likely to be granted independence as early as physically accelerated boys. Therefore, it may be anticipated that more late- than early-maturers regard adults, particularly their parents, as dominating, forcing them to do things they don't want to or preventing them from doing things they want to do (high scores in *p Dominance*). Moreover, the parental treatment these boys experience and parental refusal to grant them independent status may be interpreted as personal rejection. Hence, we predicted that more late-maturing boys would score high in *p Rejection*.

3. These feelings of being dominated and rejected may result in attitudes of rebellion against the family and in feelings of hostility. We therefore expected that more of the late-maturing group would reveal strong aggressive needs (high scores in *n Aggression*) and desires to escape from (*n Autonomy—leaving parents*), or to defy, the family (*n Autonomy—defying parents*).

4. On the basis of the data indicating that slow-maturers showed a great deal of social interest (although often of an immature kind), we hypothesized that more members of this, than of the early-maturing group would reveal strong interests in friendly, intimate interpersonal relationships (high scores in *n Affiliation*).

5. Assuming that, as Jones and Bayley (1950) suggest, the social initiative and attention-getting devices of the late-maturers are of a compensatory nature, we would expect this group to be basically dependent and to have strong needs for support from others. These should be manifest by higher scores in TAT *n Succorance* and *p Nurturance*. The latter may be considered a more indirect measure of dependence, a kind of wish-fulfilling view of the world as helpful and friendly.

6. The early-maturer, being regarded and treated as more adult, is more likely to become self-confident, and to acquire high status goals. For these reasons, we predicted that more of the physically accelerated would give evidence of high achievement goals (high scores in *n*

Achievement) and concern with personal recognition (high scores in *n Recognition*).

7. Late-maturing boys in our culture probably face more problems of personal adjustment than do their early-maturing peers. As a result of this, they may become more aware of their problems, and, as the high degree of flexibility of young adults who had been retarded in maturing suggests, more insightful. Hence we predicted that they would be more willing and able than early-maturers to face their own feelings and emotions (low scores in the TAT variable *denial of feeling*).

In summary, we attempted to test seven propositions related to differences in the personalities of early- and late-maturing boys. It was hypothesized that more late-maturers would score high in variables relating to negative self-conceptions, dependence, aggression, affiliation, rebelliousness, and feelings of being dominated and rejected. More early-maturers, on the other hand, were expected to reveal strong achievement and recognition needs, feelings of personal success, and tendencies toward denial of feelings.

PROCEDURE

The 33 seventeen-year-old male subjects of this investigation were members of the Adolescent Growth Study which included a normal sample of boys in an urban public school system. The subjects of the present investigation represented two contrasting groups, selected on the basis of their physical maturity status: 16 of them had been among the most consistently accelerated throughout the adolescent period; the other 17 had been among the most consistently retarded. All of them took the Thematic Apperception Test, which provides the basic data of this study, at age 17.

The TAT consisted of 18 pictures: nine from the Murray set which is now standard (cards 1, 5, 6, 7BM, 10, 11, 14, 15, 17); five pictures from the set generally used in 1938 when these data were collected (a man and woman seated on a park bench; a bearded old man writing in an open book; a thin, sullen, young man standing behind a well-dressed older man; a tea table and two chairs; an abstract drawing of two bearded men); and four designed especially for this investigation (the nave of a large church; a madonna and child; a dramatic view of mountains; a boy gazing at a cross which is wreathed in clouds).

The tests were administered individually. Each card was projected on a screen while the subject told a story which was recorded verbatim. Standard instructions were given for the Murray cards, and subjects were asked to describe the feelings elicited by the other four pictures. Most of the stories were brief, consisting of only one or two sentences.

As we noted earlier, each of the personality variables involved in the seven propositions was translated into a TAT scoring category. The scoring scheme involved counting the relevant needs, press, and descriptions of the heroes of the stories, the assumption being that the storyteller has identified with the hero: the hero's needs are the same as the boy's; the press that impinge upon the hero are the ones that affect the boy telling the story. A total of 20 needs, press, and descriptive categories, each defined as specifically as possible, was developed in the analysis of the protocols. A score for each subject for each TAT category was derived by counting the number of stories in which it appeared. A list of the categories used, together with brief descriptions of them, is found in Table 1.

To test the reliability of this analysis, one of the authors (PM) and another psychologist independently scored 15 complete protocols (300 stories). The percentage of interrater agreement was 90, computed by the usual formula (number of agreements divided by number of agreements plus number of disagreements).

In order to eliminate bias, the scoring used in the present study was done "blind," that is, independently of knowledge of the subject's maturational status.

RESULTS

Frequency distributions of the scores of all subjects were made for all the TAT variables. Each distribution was then dichotomized at the point which most nearly enabled the placing of half of the 33 subjects above, and half of them below, the dividing point. Subjects having scores above this point were considered high in this particular variable; those with scores below this point were considered low in this variable. Chi square tests were used to test the seven propositions, i.e., to ascertain whether or not high scores in certain TAT variables were in fact more characteristic of one group (late- or early-maturers) than of the other.

Table 1 lists the TAT variables, the number of late- and early-maturers with high scores in the variable, the chi square value obtained, and the level of significance. It should be noted that the hypotheses tested were one-sided hypotheses, while the chi square value is in terms of a two-sided hypothesis. When chi square has only one degree of freedom, the square root of chi square has a distribution which is the right hand half of a normal distribution. In order to test a one-sided hypothesis, the chi square test must be converted into the equivalent value in terms of a unit normal deviate. The levels of significance reported in Table 1 were evaluated in these terms.

Table 1 shows that, as had been predicted, more late-maturing than early-maturing boys revealed feelings of inadequacy and negative self-concepts, i.e., scored high in the TAT variable *negative characteristics*. Hence proposition 1 was confirmed. This finding is consistent with the frequently made clinical observation that retardation in physical maturation may be an important source of personal maladjustment and attitudes of inferiority.

Proposition 2 stated that more late-maturers regard their parents as highly dominating and rejecting. The evidence summarized in Table 1 substantially supported this proposition. While the difference was not statistically significant, more late- than early-maturers scored high in *p Dominance by parents* (total). There was a marked difference between the groups in the variable which involves parental domination by forcing the child to do something he does not want to do (*p Dominance by parents, forcing*). However, examination of the data with respect to the variable *p Dominance by parents (prevention)* makes it necessary to reject that part of the proposition which maintains that late-maturers are more likely to view their parents as highly restrictive of their activities.

That aspect of proposition 2 which deals with feelings of rejection was confirmed by our data. Compared with the early-maturing group, a significantly greater proportion of the late-maturers told stories in which the hero was rejected by parents or authority figures. These feelings of rejection may stem from different sources. In some cases, the parents' behavior may make it clear that they are disappointed in their physically retarded son whom they regard as immature. The boy, perceiving this attitude, may interpret it as rejection. In other cases, parental reluctance to allow the late-maturing boy to establish his independence may lead to considerable tension in the family and the boy's feelings of rejection may simply reflect the ongoing parent-child conflict.

It is possible that earlier in their teens, soon after the physical changes of adolescence became apparent, many of the early-maturing boys also experienced conflicts with their parents, arising from difficulties in establishing their independence or in handling emerging heterosexual interests. At that time they too may have felt dominated or rejected. However, by the age of 17, when these data were collected, these boys were ordinarily treated as adults and granted more freedom. Hence, they were more likely to have resolved many of their conflicts with their parents and to feel accepted and independent.

The hypothesis (part of proposition 3) that more late-matürers would be highly aggressive was rejected on the basis of the evidence given in Table 1. In fact, the differences between the two groups on all the TAT aggression variables were in the opposite direction from the prediction. High scores in the variables relating to aggression of the most overt and violent type were significantly more frequent among the early-maturers, and more members of this group also scored high in measures of milder (verbal) aggression and of total aggression. While late-maturers may experience more problems of adjustment and greater frustrations than their early-maturing peers, they apparently do not manifest greater aggressive motivation. It may be that their own feelings of inadequacy or fears of retaliation and punishment for aggression inhibit their expression of hostile feelings, even in fantasy. On the other hand, the early-maturers who feel more secure personally, and recognize their own relatively advantageous physical and social status, may feel freer to express their aggressive needs. Since aggression is a culturally stereotyped masculine trait, it seems possible that the physically accelerated, being accepted as mature and identifying readily with adult males, are more likely to acquire this characteristic. In any case, the finding that early-maturers express higher aggressive motivation during late adolescence seems consistent with Jones' finding that, as young adults, they score high on the dominance scale of the Edwards Personal Preference test (1957). Perhaps the relatively strong aggressive motivation of the early-maturer, or the mature sex-role identification it may imply, serves as a basis for the development of later qualities of leadership and persuasiveness.

As Table 1 indicates, the other aspect of prop-

Table 1: *Number of early- and late-maturers scoring high in TAT variables*

TAT Variable	Definition of Variable	High Early-Maturers	High Late-Maturers	Chi Square Value	p
Proposition 1					
Negative Characteristics	H is described in negative terms (e.g., imbecile, weakling, fanatic)	5	13	6.80	<.01
Proposition 2					
p Dominance 1	H forced by parents to do something he doesn't want to	4	8	1.73	.09
p Dominance 2	H prevented by parents from doing something he wants to	6	8	.31	>.30
p Dominance 3	Total instances of H's being forced by parents to do something and/or prevented from doing something	7	11	1.46	.11
p Rejection	H rejected, scorned, or disapproved of by parents or authorities	5	11	3.69	.03
Proposition 3					
n Aggression 1	H is aggressive in physical, asocial way	8	3	3.88	.02
n Aggression 2	H is mad at someone, argues	7	4	1.52	.10
n Aggression 3	Total of all H's aggressive actions	11	8	1.26	.10
n Autonomy 1	H leaves home	7	10	.75	.20
n Autonomy 2	H disobeys or defies parents	7	11	1.46	.11
n Autonomy 3	Total of instances in which hero leaves and/or defies his parents	3	9	4.16	.02
Proposition 4					
n Affiliation 1	H establishes good relations with his parents	8	8	.00	>.50
n Affiliation 2	H falls in love, has a romance, marries	9	14	2.66	.05
n Affiliation 3	Total instances in which H establishes and/or maintains friendly relations	8	12	1.46	.11
Proposition 5					
n Succorance	H feels helpless, seeks aid or sympathy	7	12	2.43	.06
p Nurturance 1	H is helped, encouraged, or given something by parents	5	8	.93	.18
p Nurturance 2	H is helped, encouraged, or given something by someone else (not parents)	8	14	3.88	.02
Proposition 6					
n Achievement	H attempts to attain a high goal or to do something creditable	9	10	.02	>.50
n Recognition	H seeks fame and/or high prestige status	9	8	.28	>.30
Proposition 7					
Denial of Feeling	S states that picture elicits no thoughts or feelings	9	5	2.43	.06

osition 3 was confirmed: a significantly greater proportion of late- than of early-maturers displayed strong motivations to escape from, or defy, their parents. These may be essentially aggressive reactions, stemming from feelings of parental domination and rejection, or they may reflect the late-maturers' awareness of their strife with their parents whom they perceive as blocking their drives for independence. These strong needs for escape and defiance may also be considered evidence of a generally immature way of handling parent-child conflicts. Perhaps, by the age of 17, the early-maturers have already resolved many of their conflicts with their families and/or have learned to handle these in less rebellious and in more direct and mature ways.

Proposition 4 stated that, compared with their early-maturing peers, more late-maturers would manifest strong needs for establishing close social contacts with others. While there was some confirmatory evidence, the results were not clear-cut. When all affiliative needs were considered together (score for *n Affiliation—total*), the group differences were in the predicted direction, but not statistically significant. Examination of the protocols revealed that almost all instances of affiliation concerned either parents or the opposite sex; there were very few stories involving close, friendly associations between like-sexed peers. The two major types of affiliation were scored separately. As Table 1 shows, late-maturers did not differ from early-maturers with respect to need for affiliation with parents, but a significantly greater proportion of the former group displayed strong motivation for heterosexual affiliation.

In view of the late-maturers' strong feelings of inadequacy and dependent needs (see below), it is surprising that a greater proportion of this group did not exhibit strong needs to establish and maintain close bonds with their parents. This may be due to the late-maturers' more intense conflicts with their parents at this age (17 years), their fears of being rejected and dominated by them, and their generally defiant attitudes which prevent them from admitting, even in fantasy, their strong underlying needs to form close contacts with them.

The significant difference between the groups in *n Affiliation (love, romance, marriage)* is subject to several possible interpretations. For one thing, this category may refer to general needs to establish close relations with others (with peers or adults other than parents) and not merely to desire for contact with the opposite sex. The set of stimulus cards may not have been adequate to elicit responses indicative of more general affiliative needs; hence, these were expressed through responses in the heterosexual affiliation category. If this is true, proposition 4 was confirmed, and the late-maturers' high scores in this variable indicate their greater general interest in establishing and maintaining friendly relationships.

It is also possible that the late-maturers' strong affiliative needs are actually directed only toward members of the opposite sex, i.e., that *n Affiliation (love, romance, marriage)* measures specifically heterosexual interests. Assuming that this is true, there is another plausible explanation for the discovered difference. As we saw earlier, the late-maturer may be afraid to admit that he desires close associations with his parents. He may also feel that his immaturity and poor reputational status prevent him from establishing successful social relationships with like-sexed peers. Hence, he may "displace" his affiliative needs to members of the opposite sex, who, in his fantasies, may seem more responsive.

A third possible explanation of the difference is based on Jones and Bayley's findings that the late-maturers show less overt interest in girls and are regarded as less good-looking (1950). From these data, it may be inferred that the physically retarded probably do not have successful and rewarding experiences with girls. Hence their heightened need for affiliation with the opposite sex, expressed in the TAT, may reflect their attempts to satisfy in fantasy needs which they cannot satisfy adequately in reality.

The data were generally supportive of proposition 5 which stated that late-maturers are likely to have strong underlying dependent needs. A higher proportion of this group than of their early-maturing peers scored high in *n Succorance*, the difference between the two groups approaching statistical significance ($p = .06$). Furthermore, high scores in the category involving receiving help and support from others (not including parents) (*p Nurturance—non-parents*)—an indirect measure of dependent needs—were significantly more characteristic of the physically retarded than of the physically accelerated. In view of the late-maturers' attitudes toward their parents, discussed above, it is not surprising to find that perceptions of parents as kindly and supportive (high scores in *p Nurturance—parents*) were not significantly more common in this group than in the early-maturing group.

On the basis of the data involving the TAT variables *n Achievement* and *n Recognition*, we rejected proposition 6 which stated that more early-maturers would be self-confident and have high needs for achievement and personal recognition. In our culture there is strong pressure to develop needs for achievement and personal recognition, and, according to our results, these needs and feelings may become intense regardless of—or perhaps in spite of—the child's maturational status, feelings of personal adequacy, dependency, and adjustment to parents.

Two interesting incidental findings from the TAT data seem to be consistent with the proposition that more early- than late-maturers are likely to be self-confident. Seven boys in this sample of 33 adolescents told stories in which the hero was helpful or kind to someone else (*n Nurturance*). Of this group, six were early-maturers, while only one was a late maturer ($X^2 = 2.09$, $p = .07$). Insofar as *n Nurturance* may be a measure of the storyteller's own feelings that he can accept an active, mature role, more of the accelerated group feel self-assured with respect to having attained mature status.

The other incidental finding which seems to support proposition 6 is based on responses only to card 1 of the Murray series which depicts a young boy contemplating a violin which rests on a table in front of him. Eight of the subjects spoke of the boy (the hero) as a prodigy or a genius. Of these, seven were early-maturers; only one was physically retarded ($X^2 = 5.25$, $p = .01$). If the attribution of this prestige status and accomplishment to the hero reflects the subject's own feeling that he has been an achiever, it follows that more of the physically accelerated have positive self-concepts. In view of the small number of cases involved, both of these findings must be considered tentative, but they do offer some evidence in support of proposition 6.

Proposition 7, which stated that relatively few of the physically retarded boys are unwilling or unable to face their own feelings and emotions, received some support from the TAT data summarized in Table 1. A smaller proportion of the members of this group than of the physically accelerated group specifically denied that the pictures evoked any feelings or emotions (e.g. "It doesn't make me think of anything"). While this variable may not adequately measure *denial of feeling* as a major defense mechanism, this result seems to indicate that late-maturers are more sensitive to their own feelings and more ready to admit and face them openly. Since these qualities are basic to the development of psychological insight, it may be inferred that late-maturers, as a group, are more likely to become insightful individuals.

DISCUSSION

The results of the study support the general hypothesis that, in our culture, the boy whose physical development is retarded is exposed to a sociopsychological environment which may have adverse effects on his personality development. Apparently, being in a disadvantageous competitive position in athletic activities, as well as being regarded and treated as immature by others, may lead to negative self-conceptions, heightened feelings of rejection by others, prolonged dependent needs, and rebellious attitudes toward parents. Hence, the physically retarded boy is more likely than his early-maturing peer to be personally and socially maladjusted during late adolescence. Moreover, some of his attitudes are likely to interfere with the process of identification with his parents, which is generally based on perceptions of them as warm and accepting (Payne & Mussen, 1956). This, in turn, may inhibit or delay the acquisition of mature characteristics and attitudes which are ordinarily established through identification with parents. Fortunately for the late-maturers' subsequent adjustments, they seem more willing and able to face their feelings and emotions. This may be a result of their awareness of others' attitudes toward their immaturity or their feelings of personal inadequacy and dependency.

The physically accelerated boys, on the other hand, are likely to experience environmental circumstances which are much more conducive to good psychological adjustment. Hence, their psychological picture, as reflected in their TAT stories, is much more favorable. By the time they were 17, relatively few early-maturers harbored strong feelings of inadequacy, perceived themselves as rejected or dominated by parents or authorities, or felt rebellious toward their families. As a group, they appeared to have acquired more self-confidence and had probably made stronger identifications with mature adults. Hence, they perceived themselves as more mature individuals, less dependent and in need of help, and more capable of playing an adult male role in interpersonal relationships.

These findings assume additional, probably greater, importance when they are considered in the light of Jones' findings on the early adult (age 33) adjustments of boys who had been retarded or accelerated in physical maturing (1957). It should be recalled that by this age physical differences between the two groups had practically disappeared. Certain important psychological differences were noted, however, and these were consistent with the differences at age 17, reported in the present study. For example, the responses of the early-maturing group to two paper-and-pencil tests revealed that, as young adults, they were more dominant, more able to make a good impression and more likely to be turned to for advice and reassurance; more self-controlled; and more willing and able to carry social responsibility. In short, they present a general picture of psychological maturity. Moreover, more of the early-maturers seemed to have made successful vocational adjustments. In contrast to this, when the late-maturers became adults, they tended to be highly dependent individuals who could be described, on the basis of their test responses, as tending to be rebellious, touchy, impulsive, self-indulgent, and insightful. Most of these characteristics are indicative of poor adjustment and psychological immaturity. Fewer members of this group had made good vocational adjustments.

The striking correspondence between the two descriptions of the groups, derived from different kinds of tests and collected at widely separated periods of time, lends further support to Jones' conclusion that "the adolescent handicaps and advantages associated with late- or early-maturing appear to carry over into adulthood to some extent" (1957). It seems clear that many attributes of adolescent personality (patterns of motivation, self-conceptions, and attitudes toward others) characteristic of late- and early-maturing boys are relatively stable and durable rather than situational and transitory. This may be attributable to the fact that in our culture adolescence is generally a critical and difficult period of adjustment. Within a relatively brief interval of time, the child must work out numerous complex and vitally important personal problems—e.g., adaptation to his changed biological and social status, establishment of independence, vocational adjustment. In dealing with these problems, he may acquire new behaviors and personality attributes which have broad ramifications, not only on his current adjustment, but also on his subsequent development. If the adolescent can cope with his problems without too much inner stress and turmoil, his self-esteem, feelings of adequacy, and consequently his subsequent adjustment, are likely to be enhanced. On the other hand, if his problems induce great tension and anxiety, he is likely to feel frustrated and inadequate, and, if these feelings are maintained, to adjust less satisfactorily as an adult.

Obviously, the adolescent's success or failure, as well as ease or tension, in handling his problems will be determined to a large degree by the sociopsychological forces to which he is subjected during this time, and these, as we have seen, may be significantly related to his rate of maturation. Thus, physical status during adolescence—mediated through the sociopsychological environment—may exert profound and lasting influences on personality. For this reason, many aspects of the adult's behavior and personality seem consistent with his adolescent adjustments, attitudes and motivations.

Insofar as our results permit generalization, they suggest that some important aspects of motivation, such as needs for achievement and personal recognition, are not significantly affected by maturational status. It may be that among subjects whose achievements are strongly encouraged and rewarded from very early childhood, the need to achieve becomes powerful and resistant to change even in the face of feelings of helplessness and inadequacy. The latter may inhibit the achievement-oriented overt behavior of some late-maturers, but the underlying motivation to achieve seems as strong in this group as it is among the physically accelerated.

In conclusion, it should be noted that, although rate of maturing and associated factors may affect personality development, the relationship between physical status and psychological characteristics is by no means simple. A vast number of complex, interacting factors, including rate of maturation, determine each adolescent's unique personality structure. Hence, in any specific instance, the *group* findings of the present study may not be directly applicable, for other physical, psychological, or social factors may attenuate the effects of late- or early-maturing. For example, an adolescent boy who is fundamentally secure and has warm, accepting parents and generally rewarding social relationships may not develop strong feelings of inadequacy even if he matures slowly. Analogously, the early-maturing boy who has deep feelings of insecurity, for whatever reasons, will probably not gain self-confidence simply because he matures early. In summary, in understanding any individual case, generalizations based on the data of the present study must be particularized in the light of the individual's past history and present circumstances.

SUMMARY

The present investigation was designed to test seven propositions concerning the relationship between rate of physical maturation and important aspects of personality structure, specifically, self-conceptions, underlying motivations, and basic interpersonal attitudes. The TAT protocols of 33 seventeen-year-old boys—16 who had been consistently physically accelerated throughout adolescence and 17 who had been consistently retarded—were analyzed according to a scoring schema involving 20 needs, press, and descriptive categories. The scores of early- and late-maturers in each of the categories were compared.

An earlier study (Jones & Bayley, 1950) demonstrated that late-maturing boys are more likely than their early-maturing peers to encounter a generally unfavorable sociopsychological environment. Analysis of the data of the present study indicates that this situation may have adverse effects on the personalities of the physically retarded. These boys are more likely to have negative self-conceptions, feelings of inadequacy, strong feelings of being rejected and dominated, prolonged dependency needs, and rebellious attitudes toward parents. In contrast, the early-maturing boys present a much more favorable psychological picture during adolescence. Relatively few of them felt inadequate, rejected, dominated, or rebellious toward their families. More of them appeared to be self-confident, independent, and capable of playing an adult role in interpersonal relationships. Early- and late-maturing groups did not differ significantly from each other in needs for achievement or personal recognition.

These findings make it clear that rate of physical maturing may affect personality development in crucially important ways. However, it is important to note that in any particular case the effects of early- or late-maturing may be significantly modified by the individual's psychological history and present circumstances.

REFERENCES

JONES, MARY C. The later careers of boys who were early- or late-maturing. *Child Develpm.*, 1957, *28*, 113–128.

JONES, MARY C., & BAYLEY, NANCY. Physical maturing among boys as related to behavior. *J. educ. Psychol.*, 1950, *41*, 129–148.

PAYNE, D. E., & MUSSEN, P. H. Parent-child relations and father identification among adolescent boys. *J. abnorm. soc. Psychol.*, 1956, *52*, 358–362.

TRYON, CAROLINE M. Evaluation of adolescent personality by adolescents. *Monogr. Soc. Res. Child Develpm.*, 1939, *4*, No. 4.

48

Self-Conceptions, Motivations, and Interpersonal Attitudes of Early- and Late-Maturing Girls

Mary Cover Jones and Paul Henry Mussen

Jones and Mussen also investigated the relationship between physical growth and self-perception in adolescent girls. Their discussion indicates some of the difficulties inherent in measurement in terms of discrepancies between self-report and TAT data on one hand and observation and rating by adults on the other. Furthermore, the period of the investigation, at age 17 rather than at the time when the subject was most out of step with her peers, raises questions about the interpretation of results.

"The changing body and the changing self" is a phrase associated with adolescent development. It suggests that the shaping into mature form of the childhood body pattern is accompanied by new self-concepts. These altered attitudes toward the self reflect at least in part the youth's response to his physical metamorphosis.

What "growing-up" connotes for the individual adolescent depends upon a complex of psychobiological factors. One of the most important of these is rate of physical maturation. Adolescent growth may be relatively regular and even, or it may be uneven or abrupt. The timing of puberty, in relation to social norms of the peer group, may present problems of special importance for some adolescents.

Previous reports of systematic comparisons between the behavior and personality characteristics of early- and late-maturing adolescents

---

From *Child Development*, 1958, *29*: 491–501. Reprinted with permission of the Society for Research in Child Development.

have indicated that acceleration in growth tends to carry social advantages for boys (Jones & Bayley, 1950; Mussen & Jones, 1957) but disadvantages for girls (Jones, 1949). At their peak of growth, early maturing girls are not only taller than their girl classmates but are actually taller than most of the boys in their class. They are conspicuously large at a time when physical size is not an asset for girls in our culture. Many girls consider tallness to be a physical stigma. At the end of adolescence the early-maturing are no longer taller than their age-mates, but in body proportion they tend to have a broad and stocky build, less attractive (in terms of current feminine standards) than the more slender physique of the late-maturing.

Among boys, ascendance in size and musculature is an asset because of our cultural values and the functional advantages of such a build for athletic prowess. This more favorable status is indicated in observational records for early-maturing boys. Staff members rated them as physically more attractive and better-groomed than the late-maturing, and in social situations they were more poised and matter-of-fact, and less attention-seeking (Jones & Bayley, 1950).

In contrast, both classmates (Tryon, 1939) and adult observers (Newman, 1946) saw the early-maturing girls as relatively submissive, listless or indifferent in social situations, and lacking in poise (Jones, 1949). Such girls have little influence upon the group and seldom attain a high degree of popularity, prestige or leadership.

The girls in the slower-maturing classification were seen as relatively more outgoing and more assured. They were eager, animated, peppy, and talkative. This behavior seems to be acceptable among girls since those who exhibit it are also described as confident and having leadership abilities (Jones, 1949). While the same characteristics of expressiveness are attributed to slow-growing boys, it is associated in their case more specifically (and especially in later adolescence) with show-off behavior, affectation, and tenseness (Jones & Bayley, 1950).

In accounting for these sex differences in the response to early or late puberty, we may note that although early-maturing boys have physical advantages over other boys and are socially in step with girls, the girl who develops earlier than her classmates may be temporarily isolated. H. E. Jones has expressed this as follows:

> The early-maturing girl quite naturally has interests in boys and in social usages and activities more mature than those of her chronological age group. But the males of her own age are unreceptive, for while she is physiologically a year or two out of step with the girls in her class, she is three or four years out of step with the boys—a vast and terrifying degree of developmental distance (1949, p. 78).

A study of responses to the Thematic Apperception Test, given to members of the Adolescent Growth Study when they were seniors in high school, yielded a somewhat unfavorable psychological picture for the late-maturing boys. Compared with their early-maturing peers, they showed greater evidence of negative self-concepts, prolonged dependency needs, feelings of rejection by others, rebellious attitudes toward parents, and strong affiliative needs. These findings were in agreement with evidence from other sources (Mussen & Jones, 1957).*

A similar TAT comparison of early- and late-maturing girls should be expected to show results different from those obtained for boys. Thus, it might be expected that early-maturing girls would reveal negative self-feeling and less satisfactory interpersonal attitudes.

PROCEDURE

The present study, paralleling that for boys, was designed to investigate the relationship between maturational status and self-conceptions, motivations, and interpersonal attitudes in a normal public school sample of girls. Personality assessment was made on the basis of their responses to the Thematic Apperception Test (TAT).

The 34 17-year-old girls of this investigation constitute approximately the 20 per cent at each extreme of the total sample of the Adolescent Growth Study, selected on the basis of their physical maturity status as determined by X-rays of the wrists and hands. Sixteen had been among the most consistently accelerated over a four-year period during adolescence; the other 18 were among the most consistently retarded. All of the subjects took the TAT at around age 17 when they were seniors in high school.

The TAT consisted of 18 pictures: nine from the Murray set which is now standard (cards 1, 5, 6, 7BM, 10, 11, 14, 15, 17); five pictures from the set generally used in 1938 when these data were collected (a man and woman seated on a park bench; a bearded old man writing in an open book; a thin, sullen, young man standing behind a well-dressed older man; a tea table and two chairs; an abstract drawing of two bearded men); and four cards not in the Murray series (a madonna and child, the nave of a large church, a dramatic view of mountains, a boy gazing at a cross which is wreathed in clouds).

---

* This article appears in this volume—*Editor's note.*

The tests were administered individually. Each card was projected on a screen while the subject told a story which was recorded verbatim. Standard instructions were given for the Murray cards, and subjects were asked to describe the feelings elicited by the other four pictures. Most of the stories were brief.

The scoring scheme involved counting the relevant needs, press, and descriptions of the heroes of the stories, the assumption being that the storyteller has identified with the hero; the hero's needs are the same as the girl's; the press that impinge upon the hero are the ones that affect the girl telling the story. A total of 20 needs, press, and descriptive categories, each defined as specifically as possible, was developed in the analysis of the protocols. A score for each subject for each TAT category was derived by counting the number of stories in which it appeared. Table 1 presents a list of the categories used, together with definitions of these categories.

To test the reliability of this analysis, one of the authors (PM) and another psychologist independently scored 15 complete protocols (300 stories). The percentage of interrater agreement was 90, computed by the usual formula (number of agreements divided by number of agreements plus number of disagreements).

In order to eliminate bias, the scoring used in the present study was done "blind," that is, independently of knowledge of the subject's maturational status.

RESULTS

Frequency distributions of the scores of all subjects were made for all the TAT variables. Each distribution was then dichotomized at the point which most nearly enabled the placing of half of the 34 subjects above, and half of them below, the dividing point. Subjects having scores above this point were considered high in this particular variable; those with scores below this point were considered low in this variable.

Table 1 lists the TAT variables together with the number of late- and early-maturers with high scores in each variable. The exact probabilities of obtaining these distributions of high and low scores in the two groups (or all other possible more extreme sets), calculated in accordance with Fisher's method (1938), are given in the fourth column. The last column gives the levels of significance of the differences between early- and late-maturing boys on these same variables, previously reported by the authors (1957).*

As may be seen from Table 1, early- and late-maturing boys differed from each other on many more characteristics than the two groups of girls did. The boys' groups were significantly different from each other, at the 5 per cent level or better, in six of the 20 variables scored, while the early- and late-maturing girls differ significantly in only two of the variables (*negative characteristics* and *n Recognition*). It should be noted, however, that the *direction* of the differences tended to be the same, rather than reversed, in the two sets of data. For example, the following similarities may be noted:

1. In this list of characteristics a significantly greater proportion of late-maturing girls than of early-maturing girls have high scores on *negative characteristics*. This finding is similar to that found in the comparison of early- and late-maturing boys. For girls, it is contrary to expectation.

2. The differences between early- and late-maturing girls in respect to *p Dominance* and *p Rejection* are similar to those for early- and late-maturing boys in these variables. These may be interpreted to indicate slightly poorer parent-child relationships among the late-maturing.

3. The early- and late-maturing boys differ significantly on *n Autonomy 1*, suggesting a greater tendency for the late-maturing to avoid or defy authority. The differences are in the same direction for girls, but are not significant.

4. Similar results for boys and girls in *n Succorance* may be interpreted as showing some tendency for stronger dependency needs in the late-maturing.

5. Similar results for boys and girls in *p Nur-*

---

* This article appears in this volume—*Editor's note*.

Table 1: *Number of early- and late-maturers scoring high in TAT variables*

TAT Variable (*with Definition*)	*High Early-Maturers*	*High Late-Maturers*	*p*	*p* (boys)
Negative Characteristics—H is described in negative terms (e.g., imbecile, weakling, fanatic)	3	13	.002	.01
p Dominance 1—H forced by parents to do something he doesn't want to	2	7	.08	.09
p Dominance 2—H prevented by parents from doing something he wants to	3	6	—	—
p Dominance 3—Total instances of H's being forced by parents to do something and/or prevented from doing something	4	10	.07	.11
p Rejection—H rejected, scorned, or disapproved by parents or authorities	3	7	.21	.03
n Aggression 1—H is aggressive in physical, asocial way	8	9	—	.02*
n Aggression 2—H is mad at someone, argues	12	14	—	.10*
n Aggression 3—Total of all H's aggressive actions	5	8	—	.10*
n Autonomy 1—H leaves home	8	10	—	.20
n Autonomy 2—H disobeys or defies parents	4	5	—	.11
n Autonomy 3—Total of instances in which hero leaves and/or defies his parents	5	7	—	.02
n Affiliation 1—H establishes good relations with his parents	9	8	—	—
n Affiliation 2—H falls in love, has a romance, marries	11	7	.08	.05*
n Affiliation 3—Total instances in which H establishes and/or maintains friendly relations	7	6	—	.11*
n Succorance—H feels helpless, seeks aid or sympathy	5	10	.16	.06
p Nurturance 1—H is helped, encouraged, or given something by parents	5	7	—	.18
p Nurturance 2—H is helped, encouraged, or given something by someone else (not parents)	5	10	.16	.02
n Achievement—H attempts to attain a high goal or do something creditable	7	13	.09	—
n Recognition—H seeks fame and/or high prestige status	3	12	.01	—
Denial of Feeling—*S* states that picture elicits no thoughts or feelings	10	9	—	.06

** Differences are in the opposite direction for boys as compared with girls.*

turance (significant in one variable for boys) may also be interpreted as indirect indications of stronger dependency needs in the late-maturing.

The chief differences between the sexes are as follows:

1. With respect to *n Aggression* more early- than late-maturing girls show "argumentative aggression," but the two groups of girls do not differ in physical aggression. On the other hand, more early-maturing than late-maturing boys show high degrees of both kinds of aggression.

2. On one category of *n Affiliation* (involving romantic love) higher proportions of high scores are shown for early-maturing girls as contrasted with their late-maturing peers. The differences between the early-maturing and the late-maturing boys are in the opposite direction for this category.

3. The variables *n Achievement* and *n Recognition* do not differentiate the two groups of boys. Among girls scores are higher for the late-maturing, very significantly so in the case of *n Recognition*.

4. *Denial of feeling* does not differentiate early-

and late-maturing girls but tends to yield higher scores for early-maturing boys.

DISCUSSION

The failure of the TAT data to support observational findings, especially with reference to the variable, negative characteristics, might be accounted for in a number of ways. Some writers report that in many cases thematic fantasies and manifest behavior operate independently and are even negatively related. If we assume this to be the case for our subjects, no further explanation would be needed. But there is also evidence from the literature that, for some groups, TAT findings and overt behavior may be congruent. Our data on boys are in line with this assumption, since, according to observational ratings, late-maturing boys tend to be socially disadvantaged, and, according to the TAT, personally more maladjusted.

The findings for girls are quite different, however. The early-maturing received more unfavorable ratings from both peers and adult observers on many characteristics. But in the TAT they appear to be somewhat better adjusted than their late-maturing peers. This discrepancy between observers' ratings and the picture derived from the personality tests may stem partly from the fact that the reported observational records represented an average of repeated ratings taken over a period of time (from 11 to 17 years) while the TAT stories were collected at the end of this period.

Girls who enter puberty early would be expected to have more difficulties in personal-social relations when they are out of phase with their group. However, after the peer group "catches up," these difficulties would be reduced. By the end of senior high school maturational discrepancies, and social distance due to this factor, would be less marked. It is also possible that even a slight improvement in status would bolster morale and be reflected in a projective technique designed to register attitudes and self-concepts.

There is some slight evidence of a trend toward improved social status for the early-maturing in observational ratings over the seven-year period. Twenty-five items concerned with appearance, emotional tone, social participation, responsiveness, and assurance were used in the comparison. Three of these reflected an improved status at the twelfth grade level for early-maturing girls. In two of these, "laughing" vs. "sober" and "sociable" vs. "unsociable," the accelerated girls, while still rated lower than the late-maturing, had improved sufficiently so that the differences between the two groups were no longer significant. But for one important characteristic, "popular" vs. "unpopular," the average ratings for the accelerated girls were now actually slightly higher than for the late-maturing, though the differences were not significant. This last year at high school was the only period when the early-maturing girls were rated by observers as above average in popularity.

It is conceivable that other improvements in social relationships were undetected because of the "halo effect" which, in spite of precautions, may have influenced observers who had rated these same adolescents in earlier years. It is not unlikely that if these girls of more mature status had been observed in social groups of their own choosing (presumably outside of school) the behavior picture might have been more favorable.

It may be noted that over the seven-year period the observational records received little corroboration from a self-report inventory. Although differences were not consistent in all categories, the early-maturing girls tended to score more favorably than the slow-maturing on "total adjustment," and also on family adjustment and feelings of personal adequacy. These data from the self-report inventory seem to be generally consistent with the findings from the TAT.

However, we may note that in both the inventory and the TAT the early-maturing girls appear in a somewhat better light than in their reputation scores or in ratings by adult observers. In some individual cases a favorable self-report score should not be taken at face value, in view of the tendency for some individuals to cover up or deny their deficiencies.

The only other variable which yields a significant difference between the maturity groups is the category *n Recognition,* late-maturing girls manifesting a greater desire for personal recognition. The results for *n Achievement,* though not showing significant differences, tend to support these findings. Other data for this group of girls would lead us to expect this relationship between maturity status and desire for recognition. Late-maturing girls were rated by adult observers as attaining higher prestige, showing more leadership, and having greater stimulus value than their early-maturing peers (Jones, 1949). They were also mentioned more frequently in the high school daily paper over a three-year period and were elected to more offices in extra-curricular activities. The late-maturing girls' leadership abilities, their greater social participation, and their apparent social success may have been more closely related to desires for recognition and achievement in the social sphere than to a need for affiliation.

It should be noted that, among boys, *n Achievement* and *n Recognition* were not significantly associated with rate of physical maturation. Perhaps this is due to the fact that for boys in our culture the pressures to strive for achievement and personal recognition are powerful and pervasive; hence, the boy's physical status may have little influence on his acquisition of strong achievement and recognition needs. Since these cultural pressures are undoubtedly less severe for girls, the strength of these personal needs may be more influenced by such factors as rate of physical maturation.

As we have pointed out in an earlier article (1957),* the relationship between physical status and psychological characteristics in boys is by no means simple. The evidence of the present study indicates that this relationship is even more complex in the case of girls. While the TAT analysis reported in this study suggests that early-maturing girls have fewer negative self-concepts and fewer needs for personal recognition, the results must be interpreted very cautiously. Since only two variables were found to be significantly related to physical status, it is obvious that many psychological and social factors are more important than rate of maturing in determining girls' self-concepts and personality characteristics. Furthermore, these data, considered together with the data from earlier studies on girls, suggest that the rate of maturation may affect overt behavior and covert characteristics in different—sometimes seemingly contradictory—ways.

It is also possible that, at least for girls, early- or late-maturing means different things at different stages of adolescent development. It has been proposed that since girls who enter puberty early are out of step physically with both the boys and girls in their classrooms, they tend to be socially handicapped during early adolescence. We have assumed that this would carry emotional hazards, and evidence is available from observational data and reputation measures to indicate that this is the case (Jones, 1949).

However, the accelerated girl may gain assurance from knowing that she is on the way toward a goal which is a common task for all adolescents, that of being an adult. By the end of high school, many girls in this group were beginning to feel that they had made satisfactory progress toward this goal. If, in addition to this, she can cope with the problems of this period without too much stress, her self-esteem and feelings of adequacy may be enhanced. A resulting improvement in self-concepts may be reflected in the relative infrequency of negative characteristics in TAT stories.

In conclusion, it is evident that each individual's unique personality structure is determined by a complex of interacting variables, including rate of maturation. Comments made by these subjects as young adults indicate that they were aware of a variety of surface phenomena which affected their adolescent adjustment:

> "High school is not a pleasant memory. I felt remote from my mother. If I could have talked to her, it would have helped" (a slow-maturer).
>
> "I wasn't very happy in adolescence. My father was out of work. I felt inferior outside my own circle of friends—I always aimed to please" (a very popular late-maturing girl).
>
> "I was slightly rattle-brained" (a popular late-maturing girl).
>
> "I didn't have much fun in high school. I look forward to more happiness now than I did when I was in high school. I was an ugly duckling" (a slow-maturer who ascribed many negative characteristics to the hero).
>
> "I seemed to be separated from friends in high school. I'm more outgoing now, less cautious and fearful" (accelerated girl).
>
> "I was overweight and sensitive about it—now I take things more for granted" (accelerated girl).
>
> "I had a feeling of being different when growing up" (accelerated).
>
> "I felt stupid in school" (accelerated girl).
>
> "I was very lacking in self-confidence in high school" (accelerated girl).
>
> "I'm more optimistic now. I didn't know many people in high school. I would make an effort to get on with people if I had it to do over again" (accelerated).

Feelings of inadequacy and isolation are expressed by these girls and they are attributed to lack of mental ability, financial difficulties, separation from parents, poor social status, overweight, and unattractiveness. They are about equally common among those whose maturational status was at one extreme or the other.

---

* This article appears in this volume—*Editor's note.*

It is obvious that the findings for this specific group of girls need to be particularized for each individual. These results might be modified also for girls in another geographical area or social level or in another generation. It is possible that school and community programs may be able to de-emphasize maturational status by providing an easier access to mixed social groups through classroom, extra-curricular, and recreational activities which cut across age classifications.

SUMMARY

The present study was designed to investigate the relationship between maturational status and TAT scores for a group of physically-accelerated as contrasted with a group of slow-developing girls from a normal classroom sample. The TAT protocols of 34 17-year-old girls—16 who had been consistently accelerated and 18 who had been consistently retarded—were analyzed according to a scoring scheme involving 20 needs, press, and descriptive categories.

The scores of early- and late-maturing in each of the categories were compared. Earlier reports had indicated that girls who reach puberty early are likely to be socially disadvantaged, at least until the rest of their age group "catch up" with them. It was assumed that this social disadvantage would be reflected in the TAT protocols and that differences between the two maturity groups in self-concepts, attitudes, and motivations would be found. Analysis of the data of the present study found few striking differences between the two groups of girls. However, early-maturing girls had significantly lower scores on the category *negative characteristics*, indicating more favorable self-concepts. This finding is contrary to what might have been expected on the basis of observational ratings by adults and reputational ratings by classmates. On the other hand, the TAT results are in line with scores (total adjustment, self-adequacy, family adjustment) on a self-report inventory.

Late-maturing girls have significantly higher scores on *n Recognition,* which is corroborated by data from other sources.

When the differences between early- and late-maturing girls are compared with the differences between early- and late-maturing boys (Mussen & Jones, 1957),* they are found to be in the same direction more often than in the opposite. These findings are interpreted to indicate that late-maturing adolescents of both sexes are characterized by less adequate self-concepts, slightly poorer parent-child relationships, and some tendency for stronger dependency needs.

It has been emphasized that complex psychological and cultural factors as well as maturational status contribute to personality development and that the pattern of these influences varies for each individual.

REFERENCES

FISHER, R. A. *Statistical methods for research workers.* 7th ed. Edinburgh: Oliver & Boyd, 1938.

JONES, H. E. Adolescence in our society. In *The family in a democratic society: anniversary papers of The Community Service Society of New York.* New York: Columbia University Press, 1949. Pp. 70–82.

JONES, MARY C., & BAYLEY, NANCY. Physical maturing among boys as related to behavior. *J. educ. Psychol.*, 1950, *41*, 129–148.

MUSSEN, P. H., & JONES, MARY C. Self-conceptions, motivations, and interpersonal attitudes of late- and early-maturing boys. *Child Develpm.*, 1957, *28*, 243–256.

NEWMAN, F. B. The adolescent in social groups. *Appl. Psychol. Monogr.*, 1946, *9*.

TRYON, CAROLINE M. Evaluations of adolescent personality by adolescents. *Monogr. Soc. Res. Child Develpm.*, 1939, *4*, No. 4.

---

* This article appears in this volume—*Editor's note.*

FAMILY INFLUENCES

49

The Use of the Parent Attitude Research Inventory with the Parents of Bright Academic Underachievers

Merville C. Shaw and Bert E. Dutton

The question of what factors are influential in the poor performance of some bright youngsters has plagued both the clinically oriented psychologist and the educator. Shaw and Dutton demonstrate that there is a relationship between parental attitudes and underachievement. Because this is an R-R type study, however, they are aware that any causal relationship has not been established. This study illustrates some of the problems of conducting field research and also the way researchers utilize and build on the work of their colleagues. The reader should re-examine Schaefer and Bell (15) in conjunction with this selection.

In an earlier study of academic underachievement, Shaw and McCuen (1960) demonstrated that children classified as academic underachievers on the basis of their grades in high school typically had a long history of such behavior. In males trends toward this kind of behavior were seen as early as the first grade and were generally a characteristic of the group by the third grade. Such early appearance was not a characteristic of the female groups, the trend toward underachievement not beginning until Grade 6.

The findings of this study suggested the possibility that conditions antecedent to school entrance might have at least a partial bearing on the development of underachieving behavior, and more specifically, that parent attitudes might be associated with failure to perform effectively in school. The present study represents an attempt to determine whether or not differences exist between the parents of achievers and the parents of underachievers with respect to their attitudes toward children and an attempt to determine, in a limited way, what the nature of these differences may be.

PROCEDURE

Tenth and eleventh grade students in a high school of 1,800 students constituted the pool from which subjects were drawn. The California Test of Mental Maturity (CTMM) was administered to all of the 850 students in these two grades as a part of the study. Only those who attained an IQ of 110 or over on the Language section of the test, using published norms, were considered potential subjects. Any student who performed at this minimum level or higher and who had achieved a cumulative high school grade point average (GPA) of 2.7 or below on a 4-point scale was considered an underachiever. Only academic subjects were used in computing these averages. The selection of 2.7 as a cutting point was based on the fact that it approximated the mean cumulative GPA on the tenth and eleventh grade classes. A student was considered an achiever if his cumulative high school GPA was 3.0 or higher. Table 1 reports the number of children identified in each category.

Following the identification of subjects, their parents were contacted, first by letter, then by telephone, to request their participation in the study. This participation involved coming to the school in the evening and participating in an individual interview and in the completion of several objective inventories including the Parent Attitude Research Inventory (PARI). Parents were not informed that it was a study involving academic achievement, but were told only that children in the top 25%, with respect to ability, were being studied. The time that any parent was involved ranged from 2 to 4 hours. Table 1 reports the number of mothers and fathers in each of the eight categories studied who completed all project requirements.

An examination of Table 1 reveals that cooperation was generally better from mothers than from fathers and from the parents of achievers than from the parents of underachievers. The fathers of female underachievers were the most reluctant group. A total of 151 parents completed

From *Journal of Educational Psychology*, 1962, *53*, 203–208

the necessary procedures. The size of the sample did not permit socioeconomic stratification. Inspection of data relative to occupational level of fathers indicated a tendency for underachievers to come from families lower on the occupational scale, but marked overlap between achiever and underachiever groups was also obvious.

The PARI was used as a measure of attitude. It was developed by Schaeffer and Bell (1958)* and has been used extensively. There are forms for both mothers and fathers. The Mother Form of the inventory consists of 23 factored scales of 5 items each with the items cyclically arranged on the question blank. The Father Form consists of 30 factored scales of 8 items each, with items also cyclically arranged. These scales have the advantage of covering, in an objective way, a wide range of parent attitudes and also of presenting data in a manner which makes statistical treatment easily possible. Four categories of answers ranging from strongly agree to strongly disagree are possible. In the present study, IBM answer sheets were provided for parents to respond on.

The PARI was administered to each parent following a lengthy individual interview. The subject of the interview was the particular underachieving child included in the sample, and instructions to parents with respect to completing the PARI were that they should complete it with this particular child in mind. Following the administration of the PARI to all parents who participated in the study, answer sheets were scored on each scale and mean scores were computed for each group included in the study. In addition, the three factors isolated by Zuckerman and Ribback (1958) from the Mother Form of the inventory were scored and comparisons were made between mothers of achievers and underachievers on each of the factors.

For purposes of comparison, the fathers of male achievers were compared with fathers of male underachievers, and the fathers of female achievers were compared with the fathers of female underachievers, although in the latter case the number of fathers involved is so small as to make the reliability of results open to question. The mothers of male achievers were compared with the mothers of male underachievers and the mothers of female achievers were compared with the mothers of female underachievers. This kind of breakdown for comparative purposes seemed appropriate in the light of previous studies indicating differences in the dynamics of underachieving behavior between males and females. The *F* test to determine homogeneity of variance and the *t* test of significance were carried out between appropriate groups. When *F* was significant at the .05 level or less, *t* was computed . . . .

Table 1: *Number of participating students and parents classified by grade, sex, and achievement level*

Achievement level	Students		Parents	
	Grade		Mothers	Fathers
	10	11		
Male achievers	17	14	22	21
Male underachievers	23	13	28	21
Female achievers	18	15	28	19
Female underachievers	8	7	8	4

Table 2: *PARI scales on which differences between mothers of female subjects were significant at the .05 level or less*

Scale number	Achievers' mean	Underachievers' mean	*F*	*t*
2	9.07	13.75	5.36***	2.50[a]*
5	7.21	12.50	7.92****	2.94[a]*
6	12.61	16.25	1.64	2.93****
11	11.04	14.38	1.54	2.23**
12	9.89	13.14	1.54	3.19****
16	8.75	11.00	1.94	2.15**
17	9.62	13.71	1.58	2.80****
19	9.57	14.88	1.21	4.09****
22	10.30	13.88	2.88	3.39****
23	9.56	12.13	1.19	2.05**

[a] *When F was significant at the .05 level or less, t was obtained by an approximation suggested by Cochran and Cox (1950).*

* $p = .05$.

** $.05 > p > .02$.

*** $.05 > p > .01$.

**** $p = .01$.

RESULTS

Table 2 reports the scales on which significant differences were found between the mothers of female achievers and female underachievers. Differences significant at the .05 level or less were found on Scale 2, Fostering Dependency; Scale

---

* This article appears in this volume—*Editor's note.*

Table 3: *PARI scales on which differences between mothers of male subjects were significant at the .05 level or less*

Scale number	Achievers' mean	Under-achievers' mean	F	t
3	10.55	12.54	1.26	2.48**
18	7.85	9.74	1.02	2.21*

* $.05 > p > .02$.
** $.02 > p > .01$.

Table 4: *Results obtained on mothers of female subjects from PARI scales analyzed according to Zuckerman's factors*

Factor	Achievers' mean	Under-achievers' mean	F	t
A	9.87	13.23	1.63	4.50**
B	10.83	12.50	2.44*	.14
C	17.27	17.25	1.65	.02

* $.05 > p > .02$.
** .01.

5, Martyrdom; Scale 6, Fear of Harming the Baby; Scale 11, Deification of the Parent; Scale 12, Suppression of Aggression; Scale 16, Avoidance of Communication; Scale 17, Inconsiderateness of the Husband; Scale 19, Ascendancy of the Mother; Scale 22, Acceleration of Development; and Scale 23, Dependency of the Mother. In all cases the mean scores of the mothers of female underachievers exceeded the mean of the mothers of achievers.

Table 3 summarizes scales which differentiated significantly between the mothers of male achievers and male underachievers. It will be readily seen that in this case only two scales differentiated significantly. These were Scale 3, Seclusion of the Mother; and Scale 18, Suppression of Sexuality. In both cases the means obtained by mothers of underachievers were larger.

The three factors derived by Zuckerman (1958) in a previous study were used as a basis for further analysis. Factor A, an Authoritarian Control factor, seems to measure authoritarian suppression and punitive and restricting attitudes, according to Zuckerman. Factor B, a Hostility Rejection factor, is described as measuring hostility toward children and husband and rejection of the maternal role. Factor C, a Democratic Attitude factor, is composed of the three positively worded scales and seems to measure democratic attitudes. The only difference was found between the mother of female subjects, with the mothers of underachievers scoring significantly higher on Factor A. Factors B and C did not differentiate between the two groups. Table 4 summarizes these results. When similar techniques were applied to the mothers of male subjects included in the group, no significant differences were found between the mothers of male achievers and the mothers of male underachievers on any of the three factors.

Analysis of the PARI scores obtained from the fathers of female subjects was carried out in a manner similar to that used for the mothers of our subjects. Application of F and t tests to the mean scale scores obtained by the fathers of female subjects revealed significant differences on four scales. These were Scale 7, Marital Conflict; Scale 18, Suppression of Sexuality; Scale 22, Avoidance of the Expression of Affection; and Scale 28, Change Orientation. These results are summarized in Table 5.

Analysis of the PARI scale for the fathers of males included in the study revealed only two scales which differentiated significantly between the two groups. These were Scale 18, Suppression of Sexuality; and Scale 12, Irresponsibility of the Father. These results are indicated in Table 6.

DISCUSSION

Examination of the results obtained from analysis of the data on mothers of our female subjects revealed 10 scales which differentiated significantly between the mothers of achievers and underachievers. To a greater extent than the mothers of female achievers, the mothers of our female underachievers appear to be more dependent, dominant, and in need of respect and the dependency of their children. They seem to have some fear of their own hostile impulses and cannot tolerate the aggressive behavior of their children.

Results obtained from the mothers of male subjects are not nearly so definite. The only two scales which differentiated significantly between the mothers of male achievers and underachievers were the Seclusion of the Mother and Suppression of Sexuality scales. The mothers of underachieving males appear to be both more seclusive and to suppress sexuality more strongly than the mothers of male achievers.

Analysis of the data obtained from the mothers using Zuckerman's three factors tends to be in line with the findings reported above. The mothers of female subjects showed significant differences on only one of the three factors. On Factor A, the mothers of underachieving girls exceeded significantly the scores obtained by the mothers of achieving girls. This is not at all surprising in view of the fact that the 10 scales on which the mothers of underachieving females receive significantly higher scores are all included in the 16 scales which make up Factor A. No differences were found between the two groups on Factors B and C. These results would seem to indicate that the mothers of female underachievers were dependent, dominant, and felt put upon in their roles of wives and mothers. They could not tolerate the expression of aggression on the part of their children and, at the same time, feared their own hostile impulses.

No differences were found on any of the three factors between the mothers of male achievers and underachievers. Apparently, among the subjects in our sample, negative feelings were most strongly exhibited by the mothers of female underachievers.

Results obtained from the fathers of female subjects were somewhat equivocal. The fathers of female achievers expressed significantly more strongly the feeling that marital conflict existed in their family. It is interesting to note in this regard that none of the mothers of these same children expressed this feeling to a significant degree. The fathers of underachieving females exceeded the fathers of achieving females on three scales: Suppression of Sexuality, Avoidance of the Expression of Affection, and Change Orientation. Interpretation of these findings leads to the conclusion that the suppression of overt demonstrations of emotion is accompanied in these fathers by a suppression of any behavior which to them implies sexuality. There is some basis for assuming a relationship between these two variables. The high score of these fathers on change orientation may indicate the existence of a subtle dissatisfaction with the role of husband and parent.

The fathers of male subjects differed significantly on two scales, and in both cases, the fathers of underachievers exceeded the fathers of achievers. These scales were Irresponsibility of the Father and, again, the Suppression of Sexuality. The suggestion of a basic anxiety with respect to sexual matters among the parents of underachieving children is strongly borne out. Again, there is a hint of dissatisfaction with the parental role on the part of the parents of underachievers.

It would appear that parents of underachieving children have more strongly negative attitudes toward these particular children than do the parents of achieving youngsters. The present study does not reveal definitely, however, whether these attitudes are one of the results of underachievement, or one of its causes. From a logical point of view, it is easier to build a case for such attitudes as causal rather than resultant factors. The early appearance of underachievement among males lends some support to the causal interpretation. The finding of negative self-concepts among underachievers (Shaw, Edson, & Bell, 1960) also is mildly supportive of parental attitudes as causal factors since negative attitudes on the part of parents toward a child would result in the formation of negative attitudes toward self on the part of the child. The problem of why only certain children in the same family are underachievers has not been explored, although presumably it is related to the fact that parents have different attitudes

Table 5: *PARI scales on which differences between fathers of female subjects were significant at the .05 level or less*

Scale number	Achievers' mean	Under-achievers' mean	*F*	*t*
7	24.89	19.75	2.67	2.39*
18	13.16	19.50	3.31	3.41**
22	15.21	21.00	1.29	3.29**
28	18.65	22.75	1.36	2.13*

* $.05 > p > .02$.
** $p = .01$.

Table 6: *PARI scales on which differences between fathers of male subjects were significant at the .05 level or less*

Scale number	Achievers' mean	Under-achievers' mean	*F*	*t*
12	14.94	17.75	2.44*	1.54*
18	14.69	18.40	1.09	2.35**

* $p = .05$.
** $.02 > p > .01$.

toward their various children. This is a problem which should not be too difficult to explore, from a methodological point of view. In the present study an effort was made to establish a response set relative to the *specific* child under consideration.

REFERENCES

COCHRAN, W. G. & COX, G. *Experimental designs*, New York: Wiley, 1950.

SCHAEFER, E. S., & BELL, R. Q. Development of a Parental Attitude Research Instrument. *Child Develpm.*, 1958, *29*, 339-361.

SHAW, M. C., EDSON, K., & BELL, H. M. Self-concept of bright academic underachieving students as revealed by an Adjective Check List. *Personnel guid. J.*, 1960, *39*, 193-196.

SHAW, M. C., & McCUEN, J. T. The onset of academic underachievement in bright children. *J. educ. Psychol.*, 1960, *51*, 103-108.

ZUCKERMAN, M., & RIBBACK, BEATRICE. Normative data and factor analysis on the Parental Attitude Research Instrument. *J. consult. Psychol.*, 1958, *22*, 165-171.

50

Family Socioeconomic Status and Adolescent Attitudes to Authority

Elias Tuma and Norman Livson

Another facet of the California Growth Study is represented here. In this paper the adolescent's attitudes toward authority, based upon interview, observation, and peer and teacher ratings, are examined in relation to family socioeconomic status. Tuma and Livson indicate the possible interaction of physical maturity and these two variables, but it is not investigated directly. This descriptive study might be contrasted with the studies of Ojemann and Snider (33) and Muuss (44).

This study explores the relationships between several indices of family socioeconomic status and adolescents' attitudes to authority in some common social contexts—in the home, in school, and with their friends. The emphasis here is not upon behavioral conformity or rebellion but, rather, upon the degree to which the adolescent—on a phenomenological level—accepts or rejects the rules and regulations set for him in a given situation. The hypotheses to be evaluated are very general ones: that the socioeconomic status of the family plays some part in determining the degree of conformity experienced by an adolescent and that the various components of this social status differ in the extent of their effects upon his attitudes to authority. While the specific nature of these effects—and their mediation—remains an active and current research question, there is consensus on the general statement that social classes in the United States differ in their characteristic socialization values and practices and, in turn, that children exhibit corresponding differences in their attitudes and behaviors (*see* recent review by Bronfenbrenner, 1958).

. . . In the present study the adolescent's self-characterization, gained from an intensive personal interview, is supplemented with evaluations of his behavior in a variety of situations in order to arrive at an inference concerning his underlying attitudes to authority. These attitudes are related, separately for boys and girls, to indices of the family social class at three age levels (14, 15, and 16 years) and in three interpersonal situations (at Home, in School, and with Peers).

METHOD

Sample

The sample used is a subsample of the Guidance Study group at the Institute of Human Development at the University of California in Berkeley (MacFarlane, 1938). The Guidance Study has been concerned with the investigation of the physical, mental, and personality development of a group of normal children. Observations of these children began at the age of 21 months and continued up to age 18. All of the subjects, serially selected from the birth certificate registry of Berkeley (an urban community in the San Francisco metropolitan area), were born in the years 1928 and 1929.

The Guidance Study group was, at the time

From *Child Development*. 1960. *31*: 387-399. Reprinted with permission of the Society for Research in Child Development.

that study began, a fair sampling of the Berkeley population but differed from the general United States urban population in respect to a lower rate of infant mortality, higher educational level, higher occupational level, higher percentage of home-owners, better household facilities, lower percentage of foreign-born parents, more Protestants and fewer Catholics and Jews, and, because they were young childbearing families, they had a lower per capita income.

A "core" group of 19 boys and 29 girls was selected from the total Guidance Study population as the sample for this study in order to permit study of changes over time for the same group of individuals. This "core" group comprises *all* the subjects available from the Guidance Study group for whom there were complete ratings of attitude to authority in all three of the interpersonal situations and at each of the three ages. In this total subsample of 48 families, 87 per cent of fathers and 83 per cent of mothers are native born; 81 per cent and 77 per cent, respectively, are Protestant. It should be noted that the children were reared during the depression years and that their adolescence, here under study, coincides with the span of World War II.

Measures

Measures of conformity. Each adolescent was rated on his attitude to authority in three situations (at Home, in School, and with his Peers) at ages 14, 15, and 16. These ratings were based upon a clinical evaluation of all information available on a child for a given situation at each age level. The sources of data included interviews with the adolescent and with his mother and, also, teacher ratings and sociometric appraisals from his school. The same five-point rating scale was employed for the three situations:

5. Hectic drive to conform.
4. Real urge to conform, to be accepted, to avoid friction, etc.
3. Occasional assertions of individuality but for the most part accepts rules, regulations, social standards without much wear and tear.
2. Tries to side-step rules and regulations to avoid conformity. Passive resistance or avoidance of situations where rules or regulations would have to be met.
1. Extremely resistive to rules, regulations, and authority. Extreme individualism and nonconformity.

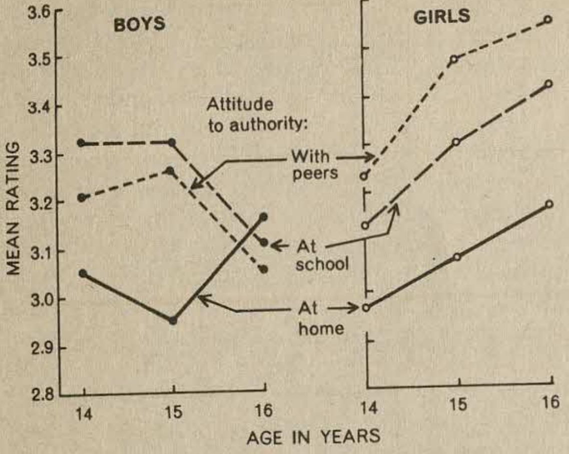
Measures of socioeconomic status. A number of socioeconomic indices available for Guidance Study families permitted the assessment of the relationships between an adolescent's conformity patterns and the socioeconomic level of his family *at the time of his birth.* An over-all measurement [denoted by SES] is provided by the California Institute of Child Welfare Index which is composed of a number of socioeconomic indices (selected by H. E. Jones). It combines four elements: (a) Berkeley Social Rating Scale (to be denoted by BSRS), which covers house exterior, neighborhood, family accommodation, living room, and special equipment; (b) total family income; (c) midparent education; and (d) occupation of father. . . . This general measure is supplemented by two additional scores intended to differentiate socioeconomic status into its "social" and "economic" components. Such a distinction seems logically tenable and is supported by the results of a factor-analytic study of all socioeconomic indices available in the Guidance Study. In this report it is found that, although a single factor may account for the bulk of covariance among 18 individual indices, it is possible to discern two empirical clusters of indices which generally separate educational and occupational items from more directly income-related items. The best measures of the former ("social") cluster are mothers' and fathers' educational level (years of schooling); for the latter ("economic") cluster the best single measurement is provided by the previously mentioned BSRS.

As can be read from Table 1, the educational level of each parent is highly related to BSRS, and all three variables show substantial correlations with SES. The median value is above .70, and all correlations are significant at the 1 per cent level or better. In view of the considerable intercorrelation of these three measures, the legitimacy of distinguishing among them depends upon their ability to yield differential correlations with the measures of attitudes toward authority.

RESULTS AND DISCUSSION

. . . Although girls show higher conformity than boys in seven out of nine comparisons, only one significant sex difference is found (age 16, Peers, $p < .01$). However, girls show a consistent tendency to increase their degree of conformity from ages 14 to 15 to 16 in all three situations,

Fig. 1: *Mean ratings of degree of conformity in three situations as a function of age.*



while boys show no clear age trend, except perhaps the suggestion that in School and with Peers they become somewhat less conforming during the age span studied. Figure 1 illustrates these age trends. We may speculate whether these reverse trends for boys and girls in conforming behavior in School and with Peers may reflect differential cultural training and values and/or the effects of differing levels and rates of physical maturation for the two sexes. Whatever the pattern of determination, there is the suggestion that, within this age range, boys are beginning to assume their "traditional" leadership and decision-making roles and girls to adopt the more conforming patterns of behavior for which they tend to be rewarded in our society.

The average ratings for both boys and girls over the three-year period fall consistently on the conformity side (above 3.0) of our scale. The grand means over the three age levels are 3.16 and 3.26 for boys and girls respectively. The distributions of ratings are generally skewed and, in a frequency distribution of all ratings, 33 per cent of adolescents are rated "4" or "5" and only 12 per cent receive ratings of "2" or "1." Over the three-year period there seems to be a considerable degree of instability in the attitude to authority in each situation. Interage correlations for boys (among the three ages over the three situations) range from .21 to .68 with a median correlation of .59. For girls, the range is .25 to .61 with a median of .43. It is possible, of course, that these low interage correlations are primarily reflections of low reliability in the ratings of conformity. However, an additional breakdown of the scores questions this interpretation: there are more changes in degree of conformity recorded for adjacent ages when the same person was responsible for both ratings than when different raters provided the data. This analysis argues that we may regard the interage correlational data as indicating that adolescent boys and girls undergo considerable fluctuations in their degree of conformity, a characteristic quite compatible with the view that the establishment of personal autonomy is a focal and wavering struggle of the adolescent period.

The interrelationships among the measures of conformity in the three situations [*see* Table 2] provide a necessary background for understanding whatever differential correlation with SES variable they may show. In their own right these interrelationships permit certain inferences concerning the situational generality of conforming behavior. At age 14 both boys and girls tend to show the same degree of conformity in School and with Peers, with the Home situation evoking more unique behavior. The tendency to show quite different attitudes toward authority in the home than outside of it is less at age 15 than at age 14 and is no longer apparent in either sex by age 16. This age change in the meaning of the Home situation, however, may be attributable to apparently opposite trends for boys and girls: for the boys, age 16 marks a period of greatly increased generality in conforming behavior; for girls, the three situations have become relatively differentiated from one another with respect to the authority attitudes they evoke.

Conformity in Relation to Socioeconomic Indices

Socioeconomic Composite (SES). Exploring the relationships between conformity of the adolescent to authority pressures at Home, in School, and with Peers and the SES of his family, we find that for boys these relationships are consistently negative (*see* Table 3). However, only two of the nine negative relationships are significant at the 5 per cent level. In contrast, there are no significant relationships between the girls' conformity and SES, nor are these relationships consistent in direction.

Evaluating the sex differences in these relationships, we find that in four of the nine comparisons the differences are significant at or

Table 1: *Intercorrelations among four socioeconomic indices* (*N = 48*)

	SES	BSRS	Education	
			Mother	Father
Socioeconomic Composite (SES)	—			
Berkeley Social Rating Scale (BSRS)	.80**	—		
Mothers' Education	.81**	.49**	—	
Fathers' Education	.71**	.35*	.70**	—

* *Significant at the .01 level.*
** *Significant at the .001 level.*

Table 2: *Intercorrelations among ratings of conformity in the home, school, and peer situations* (*19 Boys, 29 Girls*)

	Age 14			Age 15			Age 16		
	Home	*Sch.*	*Peer*	*Home*	*Sch.*	*Peer*	*Home*	*Sch.*	*Peer*
Boys:									
Home		.64	.53		.74	.51		.84	.73
School			.83			.56			.74
Girls:									
Home		.51	.53		.62	.44*		.58	.59
School			.86			.77			.63

* *Only this correlation does* not *reach the .05 level of significance.*

Table 3: *Correlations between socioeconomic status composite (SES) and conformity in home, school, and peer groups*

		ATTITUDE TO AUTHORITY IN								
	N	*Home*			*School*			*Peers*		
		14 yrs.	15 yrs.	16 yrs.	14 yrs.	15 yrs.	16 yrs.	14 yrs.	15 yrs.	16 yrs.
Boys	19	–.22	–.51*	–.38	–.38	–.42	–.33	–.37	–.44	–.51*
Girls	29	.22	.18	–.25	.26	–.02	.19	.24	–.04	.10
Significance of Sex Difference		*ns*	.02	*ns*	.04	*ns*	*ns*	.05	*ns*	.04

* *Significant at the .05 level.*

beyond the 5 per cent level. It seems fair to suggest from these data that, the lower the SES of his family, the more compliant will be the boy to authority whether it occurs in the home, in school, or in the "outside" world in general. However, before attempting an interpretation of this phenomenon on the basis of the SES composite factor alone, we will proceed to explore the relationship of conformity to its more purely "economic" and more purely "social" components.

The Berkeley Social Rating Scale (BSRS). Correlations between ratings of the family on the BSRS and the adolescent's degree of conformity remain consistently negative for boys with but a single exception; however, only one correlation achieves significance. In the case of the girls there are no significant correlations and, since the relationships fluctuate in direction, no general trend can be detected. Furthermore, there are no significant sex differences in these relationships. Since the "economic" component of the SES composite shows such low relationships, we are encouraged to expect that the "social" component (i.e., parental education) will prove more predictive.

Parents' Education in Relation to Conformity

The parents' education ranges from 5 to 19 years of schooling for both fathers and mothers. The mean score for mothers is 11.8 with a standard deviation of 3.4 years; for fathers the mean value is 12.1, the standard deviation is 3.7.

The relationships between parents' education and adolescents' attitudes to authority reproduce [in Table 4], but in a much more clear-cut fashion, those found with the SES scale. For boys, of the nine correlations with mothers' education, all are significant beyond the 5 per cent level, and four of these beyond the 1 per cent level. All are negative. With fathers' education, boys' attitudes to authority show consistently negative relationships, four of these significant at the 5 per cent level (two at the 1 per cent level). In comparison with the corresponding SES correlations, mothers' education always indicates a stronger negative relationship while fathers' education is, in general, also a better predictor of attitude to authority than is the SES composite.

In contrast with boys, there are no significant relationships between either of the parental education indices and the girls' attitudes to authority.

The data clearly admit the conclusion that an inverse relationship exists between boys' degree of conformity and the socioeconomic level of their families and that this relationship most clearly resides in the "social," rather than in the "economic," component of social status. There are a number of ways in which we may account for this general finding, and it is doubtful whether any single mediating hypothesis can encompass the full range of linkages between class and conformity which occur in the present sample. Also, although we have implicitly regarded the socioeconomic indices as continuous variables (by employing correlational analyses), the concept of social classes or strata argues for the existence of discontinuities along socioeconomic dimensions. Thus, an hypothesis put forward to account for high conformity in boys from somewhat lower social positions does not necessarily find its polar opposite in—and may even be irrelevant to—the determinants of less conformity in relatively higher classes. In this instance our sample of boys is essentially middle class (14 of the 19 families are so designated by the Warner classification; the remaining 5 boys come from upper-lower families) so that the relationships reported here largely reflect comparisons between upper-middle and lower-middle homes.

In lower-middle families we suggest that parents are more likely to imbue their sons with the importance of "respectable" behavior and also with a greater acceptance of their social status —and thus of authority. Upper-middle families, by contrast, may be more likely to value an improvement in their social positions and to implement these strivings through the manner in which they raise their children. Recalling the specific context for this investigation, it is not difficult to attribute to a highly-educated, upper-middle class mother in the 1940's a striving for upward mobility which, in part, is manifested in a need to insure, through her childrearing "philosophy" and practices, that her son achieves highly in school, in status among his peers, and in his future vocation. Under such compelling pressures reactions of both greater independence and rebellion in her son are predictable.

The formulations for boys' conformity gain some support from the literature on social class differences in parental values and practices. However, such data are notorious for their dependence upon the particular temporal and subcultural contexts in which they were gathered so that their relevance to the present findings, which themselves are tied to an atypical social period, is at best tangential. Nevertheless, we are encouraged by Duvall's (1946–1947) conclusion that working-class mothers, more than those from the middle class, stress the importance of children's obedience to their parents—and to adults in general. These data were obtained in 1943–1944, a period coinciding with our own observations. More recently, Kohn reported that ". . . middle-class parents (fathers as well as mothers) are more likely to ascribe predominant importance to the child's acting on the basis of internal standards of conduct, working-class parents to the child's compliance with parental authority" (1959, p. 341). Specifically, he reports that such traits as self-control, dependability, independence, and curiosity are more valued by middle-class parents, while working-class children are more often expected by their parents to be neat, clean, and obedient. These differences hold up for break-downs based on mothers' education, even within social classes. Coupling these observations with Bronfenbrenner's (1958) general conclusion, based upon an intensive review of studies in this area, that, in the period 1943–1953, middle-class mothers have had higher expectations for achievement for their children than did working-class mothers, we can draw some support for our interpretations of the relationship of social class variables (and particularly mothers' education) with the degree of independence shown by adolescent boys. If, indeed, more highly educated mothers are more likely to combine pressures for independence and achievement and mothers with less education to stress compliance with adult authority, then our results for boys become comprehensible.

However, since much the same class differentials in parental expectations have been found for girls, the riddle posed by their failure to show class-related variation in attitudes to authority is intensified. We can only suggest, as one contributing factor, that the expectation in the larger culture of greater conformity in girls may be equally implemented in families along the full socioeconomic range (or, at least, throughout the middle class). This is not to argue that the intentions of socialization for girls do not vary with social class; the higher-status families may be training for social poise, while lower-status families may—as we suggested for boys—value "respectability" in their daughters. We must seek elsewhere for the determinants of individual variations in conformity among girls.

Composite socioeconomic status, as measured by SES, is by itself a good predictor of boys' attitudes to authority, but more powerful indicators are the educational levels of the parents, particularly mothers' education. This is in line with the above hypotheses which assume that the mother is the primary agent mediating social class and boys' behavior. That the economic level shows such low relationships with the degree of conformity may be due to the fact that the sample has a relatively restricted socioeconomic range. Furthermore, the impact of economic factors upon attitudes toward authority may not yet be manifest by age 16. Somewhat later, when the adolescent has faced more directly the restraints on his educational and occupational mobility due to his economic level, then these factors possibly will come to bear a relationship to his conformity.

One further analysis may be of interest. Throughout this report there has been little systematic variation attributable to chronological age, per se, which is not surprising in view of the overriding importance of differential maturity rates during this period. For boys especially, the adolescent years 14 to 16 represent a time of great disparity in physical characteristics relatable to growth rate. Since physiological maturation is reflected in changes in bodily functioning (and in one's perception of one's own body) as well as in changes in one's social stimulus value (being seen less as a child and more as an adult), we were encouraged to examine the interaction between maturation, social status, and the adolescent's attitudes toward authority.

Physiological maturity status was measured, both for boys and girls, by a general (first) factor score evolved from a factor-analytic study of the interrelationships among a large number of

Table 4: *Correlations between parents' education and conformity in home, school, and peer group*

		ATTITUDE TO AUTHORITY IN								
	N	*Home*			*School*			*Peers*		
		14 yrs.	15 yrs.	16 yrs.	14 yrs.	15 yrs.	16 yrs.	14 yrs.	15 yrs.	16 yrs.
Mothers' Education:										
Boys	19	–.24	–.65**	–.49*	–.63**	.–55*	–.50*	–.65**	–.55*	–.58**
Girls	29	.32	.16	–.20	.17	.03	.17	.07	–.18	.18
Significance of Sex Difference		.08	.004	*ns*	.01	.05	.03	.01	*ns*	.01
Fathers' Education:										
Boys	19	–.29	–.58**	–.36	–.46*	–.54*	–.26	–.43	–.59**	–.37
Girls	29	.28	.15	–.08	.10	–.01	.20	.29	–.16	.24
Significance of Sex Difference		.08	.01	*ns*	.07	.07	*ns*	.02	*ns*	.05

* *Significant at the .05 level.*
** *Significant at the .01 level.*

physical development indices obtained on Guidance Study children throughout its span (Nicholson & Hanley, 1953). Dichotomizing on this factor score and on the socioeconomic composite (at their respective medians), four subgroups of boys were constituted: Early Maturing-High SES, Early Maturing-Low SES, Late Maturing-High SES, and Late Maturing-Low SES. While the sizes of these groups (5, 5, 5, and 4, respectively) precludes statistical tests of adequate power, one trend in their mean conformity scores is sufficiently consistent to merit mention: the Late Maturing-High SES boys are least conforming (or tied for lowest rank) in each of the nine possible intergroup comparisons (each situation at each age). We may speculate that the additional pressures for social achievement (which we postulated for higher SES homes) bear even more heavily upon the late-maturing boy who, in adolescence, is less well-equipped to meet the achievement standards set by his home and by his peer social groups. This isolated finding becomes more meaningful when considered in the context of some data on the personality correlates of maturation rate. Jones and Bayley (1950) conclude, from their study of the relationships between maturational status and social behavior in adolescent boys, that late maturers exhibit ". . . a 'natural' or 'childish' expressiveness in the earlier years of adolescence, and a more compensatory attention-seeking expressiveness in the later years" (p.. 137). Working with the same sample, Mussen and Jones (1957)* report that late-maturing boys show ". . . strong feelings of being rejected and dominated, prolonged dependency needs, and rebellious attitudes towards parents" (p. 255). However, they also are rated higher than are early-maturing boys on the drive for social acceptance by their peer group with the qualification, in line with the earlier study, that ". . . the late maturer's high social drives . . . are often manifested in childish, affected, attention-getting social techniques" (Mussen & Jones, 1958, p. 66). The gist of these findings suggests that the low conformity in School and with Peers of the Late Maturing-High SES boy in the present study may stem primarily from the behavioral unconventionality of the social techniques he employs in order to impress and win over his schoolmates and friends while, in the Home situation, low conformity denotes true rebellion against parents. The Late Maturing-Low SES boy, by contrast, is—on the average—the most conforming of the four groups. We may wonder whether this is due to a greater acceptance of, or resignation to, his physical retardation and whether, in turn, his conformity is an expediency made possible by the lesser achievement pressures of his lower-middle-class home.

Replicating this analysis for girls, for whom the effects of differential maturation should be less by this age due to their earlier maturation, we find only that Early Maturing-High SES girls are most conforming of the four groups in School and with Peers: at all three ages they show the highest mean ratings in these two situations. One possible interpretation is that more mature girls from High SES families accept earlier and more rapidly internalize the standards for "good" behavior put forward by their socially superior (and perhaps more socially striving) parents.

The suggestions arising from this additional analysis have, as is evident, only very fragile support from our data. At the very least, however, they do indicate the potential usefulness of regarding the adolescent's physical status as a relevant mediator of the effects of his family's social status and its accompanying values upon his attitudes toward authority, both in the home and in his social environment.

SUMMARY

Attitudes to authority, ranging from conformity to rebelliousness, were evaluated in three interpersonal situations (at Home, in School, and with Peers) for the same sample of adolescent boys and girls at ages 14, 15, and 16 years. These inferential ratings, based on interviews with the subjects and their parents and on data provided by teachers and classmates, were analyzed for sex differences (a slight tendency toward greater acceptance of authority in girls), for generality among the three ratings at a given age (moderately positive intercorrelations), and for stability of a given attitude over the three age levels (considerable fluctuation).

Most striking is the consistently negative relationship between degree of conformity experienced by the male adolescent (in all situations and all ages) and the socioeconomic status of his family (measured separately by a composite index, by the Berkeley Social Rating Scale, and by mothers' and fathers' educational levels). No consistent trend is apparent for girls. Mothers' education is the single, most powerful predictor of boys' attitudes to authority, yielding significant negative correlations in eight of nine instances. There is a suggestion, for both sexes, of variation in degree of conformity relatable to an interaction between socioeconomic factors and the physical maturity level of the adolescent. The results are evaluated in the context of reported social class differences in parental values and practices and their corresponding socialization goals.

---

* This article appears in this volume—*Editor's note.*

REFERENCES

BRONFENBRENNER, U. Socialization and social class through time and space. In Eleanor E. Maccoby *et al.* (Eds.), *Readings in social psychology.* New York: Holt, 1958.

DUVALL, EVELYN M. Conception of parenthood. *Amer. J. Sociol.*, 1946–1947, *52*, 190–192.

JONES, MARY C., & BAYLEY, NANCY. Physical maturing among boys as related to behavior. *J. educ. Psychol.*, 1950, *41*, 129–148.

KOHN, M. L. Social class and parental values. *Amer. J. Sociol.*, 1959, *64*, 337–351.

MACFARLANE, JEAN W. Studies in child guidance: I. Methodology of data collection and organization. *Monogr. Soc. Res. Child Develpm.*, 1938, *3.*

MUSSEN, P. H., & JONES, MARY C. Self-conceptions, motivations, and interpersonal attitudes of late- and early-maturing boys. *Child Develpm.*, 1957, *28*, 243–256.

MUSSEN, P. H., & JONES, MARY C. The behavior-inferred motivations of late- and early-maturing boys. *Child Develpm.*, 1958, *29*, 61–67.

NICOLSON, ARLINE B., & HANLEY, C. Indices of physical maturity: derivation and interrelationships. *Child Develpm.*, 1953, *24*, 4–38.

THE PEER SOCIETY

51

The Adolescent Culture

James S. Coleman

Coleman, a sociologist, conducted a large-scale survey of adolescent and parental attitudes in several midwestern communities, and this selection is part of the larger report to the U. S. Office of Education, which sponsored the research. Although the findings are not startling, the study shows how survey techniques are utilized to gather data on large numbers of people in various geographical settings. It also offers an interesting way of graphically presenting data for trichotomous items.

The increasing importance of the adolescent culture has a number of implications important for educational theory and practice. These implications stem from the simple fact that adolescents are not looking to the adult community for their social rewards, but to each other. To be sure, parents and parental desires are of great importance to children in a long-range sense, but for their everyday activities, in school and out, it is their peers whose approval, admiration, and respect they attempt to win. As a consequence, the old "levers" by which children are motivated, the approval or disapproval of parents and teachers, are less efficient. Educational practice has never ceased to use these levers, for it has recognized (as sometimes educational theory has failed to do) that if a child is to learn, and to work at learning, he needs both a challenge and meaningful social rewards when he meets the challenge.

So long as those social rewards could be directly provided by adults, there was little need to be explicit about them in educational theory, for

---

From *Social Climates in High Schools, Cooperative Research Monograph No. 4* by J. S. Coleman (Washington, D.C.: U.S. Department of Health, Education, and Welfare), pp. 9–29. This monograph is based on research supported by the Cooperative Research Program, U.S. Office of Education.

they were naturally provided by the very process of interaction between parent and child or student and teacher. To be sure, these rewards were often distributed in ways which reinforced the stratification system and took away the lower-class child's meager chance for equality;[1] and, as Allison Davis has shown, the middle-class backgrounds of teachers often make them unable to hold out reasonable rewards for reasonable achievement to lower-class children.[2] Yet the situation was fundamentally simpler than it is today, because teachers and parents had direct control over levers they could apply to motivate children. Now the levers are other children themselves, and adults must come to know either how to control these levers, or else how to break down the adolescent culture, thus reestablishing control by the old levers.

This latter solution might be exceedingly difficult, for it flies in the face of large-scale social changes, and undoubtedly would require a reorganization of work and community which is hardly in the offing. The major thesis of this research is that it is possible to take the other approach, to learn how to control the adolescent community as a community, and to use it to further the ends of education.

The first step in doing this is to examine a number of adolescent communities themselves, in order to discover just what the value systems are. What are the bases on which adolescents give approval to one another or withhold it; what are the grounds for becoming a member of the "inner core" or "leading crowd"; what are the things which make a boy or girl popular, admired, and imitated by his fellows? There are differences from community to community and from school to school, some of which will be examined in detail later; but there are also similarities, which make it worthwhile investigating the values of the general adolescent culture and the possible effects of these values on children. The differences indicate that it is not hopeless for adults to attempt to modify the adolescent cultures. The similarities among different schools suggest there are some general elements in the role to which adolescents are relegated by the adult society.

VALUES AND ATTITUDES OF TEENAGERS

Because adolescents live so much in a world of their own, which they do not share with older people, adults remain uninformed about the way teenagers spend their time, the things that are important to them, and the things that friends have in common. Several questions which were asked in the study give a picture of their patterns of activities and interests.[3] One question asked of the boys was:

> "If you could be any of these things you wanted, which would you want to be?"

The replies were (figures represent per cent; No-answers were excluded):

Boys most wanting to be:	Fall (*N* = 3,892)	Spring (*N* = 3,746)
Jet pilot	31.6	31.3
Nationally famous athlete	37.3	36.9
Missionary	5.7	5.9
Atomic scientist	25.6	25.9

These results strikingly show the way in which the adolescent culture departs from the educational norm: The nationally famous athlete leads, the atomic scientist is a poor third after the jet pilot, and the missionary is almost neglected. The shifts over the period of the school year are little, if any. This result of course differs from school to school, but it is important to note that in *every school* the nationally famous athlete and the jet pilot were first and second in popularity. These adolescent attitudes seem to reflect the dominant mass-media themes and heroes far more than the heroes their teachers and parents would have them follow. It may be that the most impressive contacts adolescents have with the adult culture are through the distorted lens of the mass media and its "newsworthy" events, just as the most impressive contacts many adults have with adolescents are through the mass media, with their focus upon newsworthy behavior.

In response to the same question, with the following alternatives, the girls gave responses slightly more in line with traditional values of service (figures represent per cent):

---

[1] This is graphically illustrated by A. B. Hollingshead in *Elmtown's Youth* (New York: John Wiley and Sons, 1949), a study which shows the various mechanisms by which the school reinforces the class structure. Because the present study includes "Elmtown" as one of its ten schools, it will be possible to show the great cross-cutting of social class lines in Elmtown, compared to 1941.

[2] Allison Davis, *Father of the Man: How Your Child Gets His Personality*, Boston: Houghton Mifflin Co., 1947.

[3] A more complete report is given in *The Adolescent Society*, published by The Free Press (Glencoe, Illinois), 1961.

Girls most wanting to be:	Fall (N = 4,057)	Spring (N = 3,922)
Actress or artist	18.4	19.8
Nurse	29.2	26.0
Model	32.0	33.5
Schoolteacher	20.6	20.6

Yet the glamorous model was first in popularity and schoolteachers third, with the nurse between. And over the period of the academic year, there was a slight *increase* for the glamorous model and "actress or artist," while the nurse *decreased* and the teacher remained constant. . . .

Another attitude question focused more sharply on the values current in the schools themselves. It asked a boy or girl how he or she would like most to be *remembered* in school, a matter which presumably reflects in part his own interests and in part the things that are important to his fellows. The question was:

> If you could be remembered here at school for one of the three things below, which one would you want it to be?

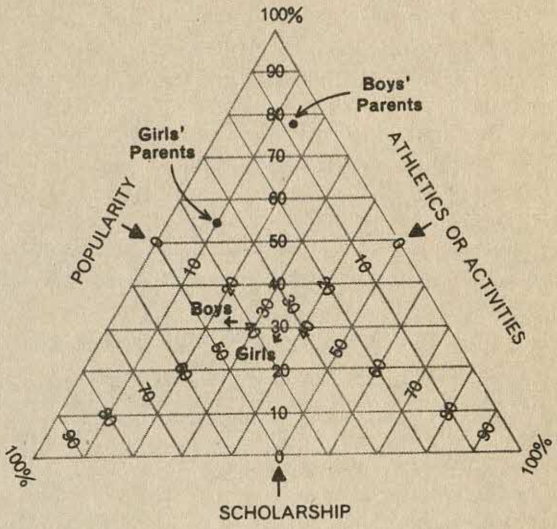
Boys:
- Brilliant student
- Athletic star
- Most popular

Girls:
- Brilliant student
- Leader in activities
- Most popular

The responses for the fall and spring questionnaires are shown in Figure 1, for both boys and girls. The three axes of the triangle indicate athletics (boys) or activities (girls), scholarship, and popularity. The more students giving a particular response, the closer the point is to the 100-per cent end of that axis. Thus a school in which 100 per cent of the boys responded "brilliant student" would be represented by a point at the upper vertex, a school in which 100 per cent responded "athletic star" would be at the left vertex, and a school in which 100 per cent responded "most popular" would be at the right vertex.

Graphs like this will be used at several succeeding points to "locate" different schools and to trace their course over the period of the academic year. In order to give some sense of the relation of these graphs to the more usual presentation of percentages, the percentages (No-answers excluded) are tabulated as follows for this first use of the triangular graphs.

Fig. 1: *How boys and girls would like to be remembered at school, and how their parents would like them to be remembered: Relative importance of scholarship, popularity, and (for boys) athletics or (for girls) activities.*



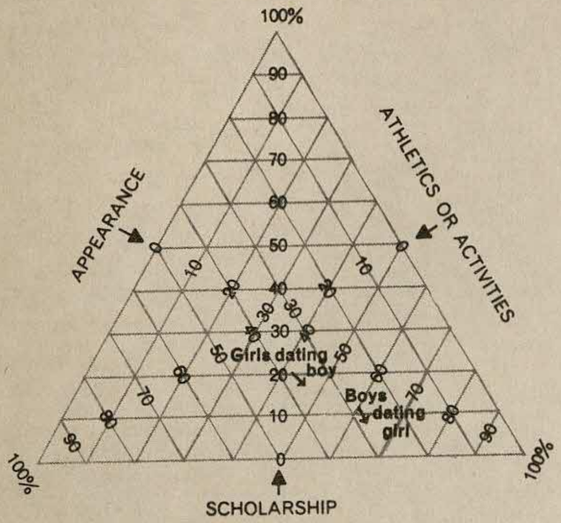
The way students want to be remembered in school:	Boys (%) Fall (N = 3,696)	Spring (N = 3,690)
Brilliant student	31.3	31.5
Athletic star	43.6	45.1
Most popular	25.0	23.4

The way students want to be remembered in school	Girls (%) Fall (N = 3,955)	Spring (N = 3,876)
Brilliant student	28.8	27.9
Activities leader	36.1	37.8
Most popular	35.2	34.2

In Figure 1, the changes from fall to spring are shown by an arrow. For boys, the athletic star's image is not only more attractive at the beginning of the school year, but the boys even move slightly more in that direction, at the expense of the popularity image, over the period of the school year.

The girls are somewhat similar: at the beginning of the school year, the "activities leader" and "most popular" are about equally attractive

Fig. 2: *Boys' and girls' choice for a date: Relative importance of scholarship (best student), appearance (best looking), and athletics (star athlete) or activities (cheerleader).*



images, both more popular than the brilliant student. By spring, the activities-leader image has gained slightly in attractiveness, at the expense of both the brilliant student and the most popular. These shifts are quite small, and there are differences from school to school . . . . Nevertheless, the point is clear: the image of athletic star is most attractive for boys and the image of "activities leader" and "most popular" are more attractive to girls than "brilliant student."

Athletics and Activities

The importance of athletics in these cultures is striking, particularly when we realize that the school as an institution is designed to focus attention on studies, and presumably upon the brilliant student.

A further view, which gives some suggestion as to why the brilliant-student image is not more popular, is given by a parallel question—not how would they like to be remembered, but whom would they like to *date*:

> Suppose you had a chance to go out with either a cheerleader, or a girl who is the best student in class, or the best looking girl in class. Which one would you rather go out with?
>
> Cheerleader
> Best student
> Best looking

> Suppose you had a chance to go out with either a star athlete, or a boy who is the best student in class, or the best looking boy in class. Which one would you rather go out with?
>
> Star athlete
> Best student
> Best looking

The responses of the boys and of the girls were as indicated in Figure 2. Perhaps the most striking result shown in the graph is the fact that the brilliant girl student fares so poorly as a date for boys. Less striking, but still evident, is the poor showing of the image of "brilliant boy student" as a date for girls. For both boys and girls, the movement over the school year is *away* from the brilliant student as someone to date. The best looking boy or girl gains at the expense of the best student, both for the boys and for the girls.

These values held by the adolescent community are fairly well known, at least by adults who are in contact with teenagers. They sometimes manifest themselves strikingly, as in the following query taken from an adolescent advice column in a church magazine:

> *Question:* If a boy isn't an athlete he isn't anything. I've been dating a girl. Last night she stood me up and went out with a football player. I get good grades and I have brains. That athlete is dumb. Why should he rate higher than I do?
>
> —A. G.

> *Answer:* Probably it isn't fair, but in high school athletes usually outrate nonathletes. Is there any sport in which you could excel? Or might you be elected to a student-body office? After graduation you'll find the really dumb athletes fade out of the scene. Among adults brains are more respected than brawn.[4]

The impact which such values have upon the way boys and girls focus their energies is not a light one. Many parents deplore these effects, as they see their sons caught up in athletics or their daughters pining to be cheerleaders or other

---

[4] *Together*, the Methodist Church Magazine. Feb. 1959, p. 45.

"important" figures. The desires of these children's parents for them is suggested by the parents' responses to a question similar to the "be remembered" question asked of the students:

> If your son or daughter could be outstanding in high school in one of the three things listed below, which one would you want it to be?

Parents of boys
- Brilliant student
- Athletic star
- Most popular

Parents of girls
- Brilliant student
- Leader in activities
- Most popular

Parental Values

The results are shown for parents of boys and parents of girls in Figure 1, imposed upon the previous graph which showed the way their sons and daughters wanted to be remembered. The sharp contrast of these values with those expressed by the students throws into relief again the divergence of the parental values and those of the adolescent subculture.

To be sure, these values may not be the ones they express day by day to their children. Indeed, the disparity between professed parental values on the one hand, and the values they express in their actions, on the other, may be very great. Replies to a set of questions asked of boys and girls suggest such a discrepancy.

> Bill (Ann) was doing well in biology class, because he (she) had a hobby of collecting and identifying insects. One day his (her) biology instructor asked Bill (Ann) if he (she) would act as the assistant in the class. If something like this happened to you, would it be something that would make your parents proud of you, or wouldn't they care?

	Per cent of boys (N = 3,831)	Per cent of girls (N = 3,956)
Both would be very proud of me	60.2	63.5
They might feel a little proud Mother would be proud, father wouldn't care Father would be proud, mother wouldn't care	31.6	30.9
They wouldn't care	5.3	4.6
No answer	2.9	1.1

> What if a different situation occurred—you made the basketball team (became a cheerleader), much to your surprise. Would that make your parents proud of you, or wouldn't they care?

	Per cent of boys (N = 3,831)	Per cent of girls (N = 3,956)
Both would be very proud of me	68.2	77.0
They might feel a little proud Mother would be proud, father wouldn't care Father would be proud, mother wouldn't care	23.8	18.9
They wouldn't care	5.6	3.4
No answer	2.5	0.8

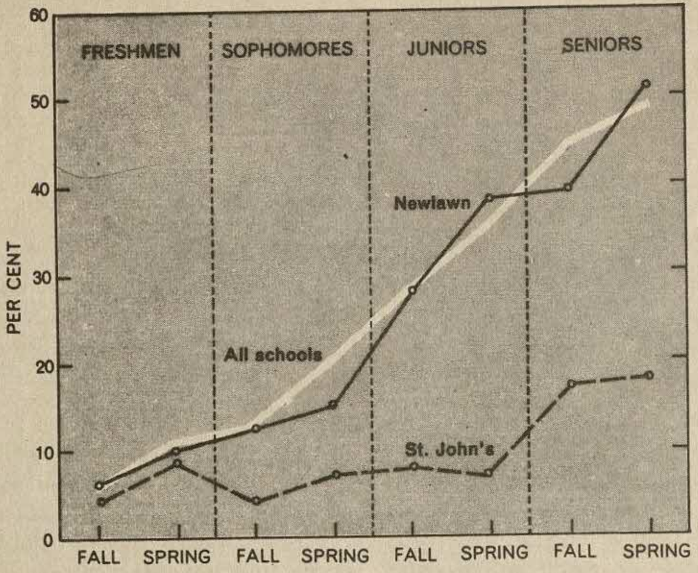
According to the boys' responses, their parents would be considerably more proud of them for making the basketball team than for being chosen biology assistant. The girls' responses indicate even more strongly that their parents would be prouder of them for becoming a cheerleader than for being chosen biology assistant.

These responses are at great variance with those the parents themselves gave, for the overwhelming majority of parents said they wanted their son to be a brilliant student in school, and a smaller majority, but still a majority, of parents said they wanted their daughters to be outstanding as a brilliant student.

Why this discrepancy? Is someone not being honest? Are the parents' professed values and expressed values as greatly at variance as they seem—or do boys and girls see their parents' values as they want to see them?

Perhaps neither is the case. It may be that most parents hold academic achievement as an ideal for their children in school, just as they expressed in their questionnaire, and as shown in Figure 1; but parents also want their children to be successful in the things that "count" in the school, that is, the things that count in the eyes of other adolescents. And parents know what things count. Being a biology assistant counts far less in the adolescent culture than does making the basketball team for boys, or making the cheerleading squad for girls. Thus in this peculiar fashion, even the rewards a child gains from his parents may help reinforce the values of the adolescent culture—not because his parents hold those same values, but because parents want to see their children successful and looked up to by the other children at school.

Fig. 3: *Per cent of high school boys owning cars in Newlawn, St. John's, and all 10 schools of the study.*



The Importance of Cars

Cars are important to a teenager. Without a car, a boy must be chauffered to movies, sports events, and, most embarrassing of all, to dates. Just as cars have become more important for adults, they have become more important for their adolescent children—as the parents of adolescent children will quickly attest. Consequently, as he reaches the age when he can possess a driver's license (16 in most States, including Illinois, in which all these schools are located), a boy's pressure on parents for permission to have a car of his own becomes extremely great. Parents find it hard not to give in, at least by letting him buy a car if he can pay for it. The great jump in car owning in the sophomore and junior years is shown in Figure 3, which shows in the fall (just after school had started) and in the spring, at the end of the school year, how car owning changes from the freshman to the senior years.

The increasing importance of cars is in part due to changes in the ecology of our lives with the blossoming of the suburbs. Boys who would have lived in the city, where a car is not important for transportation, now live in the suburbs, where it is important. A comparison of increase in car ownership from freshmen to seniors in St. John's High, in the center of Chicago, with Newlawn, in a suburb, shows this well (Figure 3). These two schools contain boys with almost identical family backgrounds (first and second generation East European immigrants, working class, predominantly Catholic); yet the car ownership increases sharply over the 4 years in Newlawn, and remains at a low level among the boys of St. John's High.

As seen in [Table 1], there is a far greater prevalence of cars among teenagers in small town and country schools than in city or suburban schools in this study. When asked if most of the boys in the senior class had cars, one boy from the smallest school, Farmdale, replied that they did because many of them lived in the country, where it was a matter of necessity.

This is certainly true, if they must have independence. However, necessity is not the whole answer; even the nonfarmers' sons in the small towns more often have a car than do their city and suburban counterparts, and the school buses run for the farmers' sons, who ride them in earlier grades. In small towns, a car is evidently more a part of the adolescent culture than it is in suburbs and cities. Yet in all the schools except St. John's High, in the center of the city, many of the students own cars by the end of their senior year.

The general level of economic prosperity accounts in part for the high ownership of cars

Table 1: *The importance of cars*

School (Approximate town size)	*Per cent owning a car*	*Number of boys (seniors)*
Farmdale (1,000)	81.3	16
Marketville (4,000)	56.4	39
Elmtown (7,000)	42.2	45
Maple Grove (6,000)	68.4	38
Green Junction (5,000)	63.8	58
Newlawn (9,000—suburb)	50.8	59
Executive Heights (17,000—suburb)	42.2	154
Millburg (25,000)	36.6	93
Midcity (110,000)	46.2	221
St. John's (3,600,000)	17.5	102

among teenagers. But this only provides the opportunity; the frequency and speed with which this opportunity is seized when a boy or girl comes of driving age indicates just how important a car is to most of the teenagers. If a boy has a car or a friend with a car, then there is no great problem in dating and in going to the local hangout; but if he doesn't, he is still a child, for he is able to date a girl only with difficulty, since public transportation is poor or nonexistent in most suburbs and small towns and is becoming poorer in cities.

LEADING CROWDS IN THE SCHOOLS

Just what does it take to "rate" in these schools, with one's own sex, and with the other sex? What does it take to be in the "leading crowd" in school? This question, of course, presumes something to which some might object: that there *is* a leading crowd in the school. And, to be sure, when students were asked such a question, some, particularly in the smallest school, did object to the idea that there was a leading crowd. Yet this kind of objection is in large part answered by one of the boys in another small school, Maple Grove, in a group interview, when his friend denied that there was any leading crowd at all in the school, and he responded: "You don't see it because you're in it." Another boy in the same school had this to say in an interview:

Q. What are some of the groups in school?

A. You mean like cliques? Well, there's about two cliques. There's one that's these girls and boys—let's see, there's ——— ——— ———. I'm in it, but as far as I'm concerned, I'm not crazy about being in it. I tell you, it wasn't any of my doing, because I'm always for the underdog myself. But I'd rather be with a bunch like that, you know, than have them against me. So I just go along with them.

Q. What's the other clique?

A. Well, I don't know too much about it, it's just another clique.

Q. Kind of an underdog clique?

A. Sort of.

Q. Who are some of the kids in it?

A. Oh—I couldn't tell you. I know, but I just can't think of their names.

Q. How do you get in the top clique?

A. Well, I'll tell you, like when I came over here, I had played football over at ———. I was pretty well known by all the kids before I came over. And when I came there was ——— always picking on kids. He hit this little kid one day, and I told him that if I ever saw him do anything to another little kid that I'd bust him. So one day down in the locker room he slammed this kid against the locker, so I went over and hit him a couple times, knocked him down. And a lot of the kids liked me for doing that, and I got on the good side of two or three teachers.

Q. What are the differences between these two cliques?

A. Well, I'll tell you, I don't like this top clique, myself. Just to be honest with you, they're all scared of me, because I won't take anything off of them, and they know it. I've had a run-in with this one girl, she really thinks she's big stuff. And I don't like her at all, we don't get along, and she knows it and I know it, and they don't say nothing. But a lot of them

> in the big clique, they're my friends. I get along with them real good, and then I try to be real nice to the underdogs, the kids that haven't got—not quite as lucky—they haven't got as much money. They have a hard time; maybe they don't look as sharp as some of the others.
>
> Q. What are the main interests of the top clique?
>
> A. Just to run everything, to be the big deal.
>
> Q. Are most of the boys in athletics?
>
> A. Yeah—you couldn't really say that in this town, though. The really good athletes, a couple of them may be in the clique—the clique's a funny thing, it's just who they want to be in it. They don't want to have anybody in there they think might give them trouble. They want to rule the roost.
>
> Q. Do most of them have fathers that have good jobs, are they well-to-do?
>
> A. Most of them. They come from families that have money.
>
> Q. Would this be the main thing that divides the top clique from the others?
>
> A. Could be, very easily, it sure could.
>
> Q. What does this underdog clique have in common?
>
> A. Well, you might say they just stick together, for self-protection. And they do things together.
>
> Q. And there are both boys and girls in the cliques?
>
> A. Yeah.
>
> Q. And they all go around together?
>
> A. Um-hm.
>
> Q. In a party, would———
>
> A. Now there you go. The big-deal clique, that's all that's there. None of the underdogs are there at all. They won't invite———I've got some graduation pictures here now, and we can get some names from that. Now she's not in the clique.
>
> Q. What's her name?
>
> A. Now this girl here, she's Joyce. She's real sweet, not very sharp-looking or nothing, and I sat by her in home room. And I talked with her and stuff like that, and she really—I mean, I think she thinks a lot of me. And this is a girl who sort of sticks to herself, more or less. She's a very nice girl, but I don't think she's in the clique. And she's not in the clique and neither is———.
>
> Q. And they aren't in the other clique either?
>
> A. No. Well, yeah, they're probably—I'll tell you, some of them—see, the big clique rules the roost, and this underdog clique, you might say, is just there to give the other one a little competition, just to know there is another one. And a lot of them are not in it. I was just automatically put in it, you might say. I didn't ask to be in it or nothing.

This account of the leading crowd in one school gives a vivid picture of how such crowds function. This is not to say, of course, that the leading crowd in every school functions in just the same way. Most interviews in other schools suggested a somewhat less closed circle than in this school, yet one which is not greatly different.

In every school, most students saw a leading crowd, and were willing to say what it took to get in it. This should not be surprising, for every adult community has its leading crowd, though adults are less often in such close and compelling communities. Yet laymen and educators as well are often blind to the fact that the teenagers in a high school *do* constitute a community, one which *does* have a leading crowd. Consequently adult concern tends to be with questions of better ways to teach "the child" viewed as an isolated entity—whether it is the "gifted child" or the "backward child."

How to Get In

In order to see more systematically what it takes to be in the leading crowd, every student was asked in the questionnaire:

> What does it take to get into the leading crowd in this school?

The major categories of response to this question about the leading crowd are shown in Figure 4. Consider first the girls' responses. Most striking is the great importance of "having a good personality." Not only is this mentioned most often in the overall study, but it is mentioned most often in seven of the nine schools.

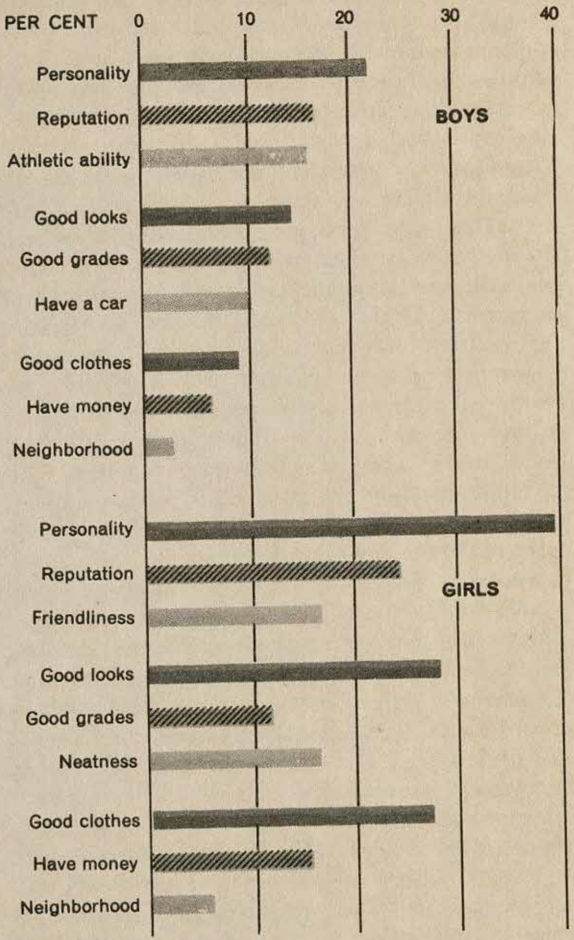
The importance of having a good personality, or, what is a little different, "being friendly" or "being nice to the other kids," in these adolescent cultures is something which adults often fail to realize. Adults often forget how person-oriented children are: they have not yet moved into the world of cold impersonality in which many adults live. This is probably due to the limits on their range of contacts; for in the limited world of grade school, a boy or girl *can* respond to his classmates as persons, with a sincerity which becomes impossible as one's range of contacts grows. One of the transitions for some children comes, in fact, as they enter high school and find that they move from classroom to classroom and have different classmates in each class.

After a good personality come a wide range of attributes and activities. A flavor of this is indicated by the collection of responses listed below—some hostile to the leading crowd (and in their hostility, often seeing it as immoral), others friendly to it (and in their friendliness attributing positive virtues to it).

> What does it take to get into the leading crowd in this school?
>
> "Wear just the right things, nice hair, good grooming and have a wholesome personality."
>
> "Money, clothes, flashy appearance, date older boys, fairly good grades."
>
> "Be a sex fiend—dress real sharp—have own car and money—smoke and drink—go steady with a popular boy."
>
> "Have pleasant personality, good manners, dress nicely, be clean, don't swear, be loads of fun."
>
> "A nice personality, dress nice without overdoing it."
>
> "Hang out at ———'s. Don't be too smart. Flirt with boys. Be cooperative on dates."

Among these various attributes, the graph shows "good looks," phrased in some fashion, to be second to "personality" in frequency for the girls. Having nice clothes, or being well-dressed, is the third most frequent item mentioned. What it means to be well-dressed differs sharply in a well-to-do suburb and in a working-class school, of course. Nevertheless, whether it is the number of cashmere sweaters a girl owns, or simply having clean and attractive dresses, the matter of having good clothes is an important one in the value systems to which these girls pay heed. This importance of clothes appears to derive in part from the fact that clothes symbolize family status. But in part also, it appears to derive from the same source that gives importance to "good looks": these items are crucial in making a girl attractive to boys. Thus in this respect the values of the girls' culture are molded by the presence of boys—and by the fact that success with boys is itself of overriding importance.

Fig. 4: *Major criteria for membership in the leading crowds of 9 schools of the study: Per cent of boys and girls mentioning each.*



Another element in the constellation of attributes required if one is to be in the leading crowd is indicated by the class of responses labeled "having a good reputation," which was fourth in number of times mentioned. In all these schools, this item was often mentioned, though in each school, some saw the leading crowd as composed of girls with bad reputations and immoral habits; and in some schools the leading crowd was less interested in a good reputation than in others.

A girl's reputation is crucial among adolescents. A girl is caught in a dilemma posed by the importance of good looks on the one hand, and a good reputation on the other. A girl must be popular with the boys, says the culture, but in doing so she must maintain her reputation. In some schools, the limits defining a good reputation are stricter than in others; but in all the schools, the limits are there to define what is "good" and what is "bad." The definitions are partly based on behavior with boys, but they also include drinking, smoking, and other less tangible matters—something about the way a girl handles herself, quite apart from what she actually *does.*

It is not such an easy matter for a girl to get and keep a good reputation, particularly if her mother is permissive in letting her date whom she likes as a freshman or sophomore. Junior and senior boys often date freshman and sophomore girls, sometimes with good intentions and sometimes not. One senior boy in Green Junction, in commenting upon the "wildness" of the leading girls in his class, explained it by saying that when his class was in the eighth grade, it was forced to go to school in the high school building because of a classroom shortage. A number of the girls in the class, he explained, had begun dating boys in the upper classes of high school. This, it seemed to him, was where the problem began.

Another criterion by which a girl gets into the leading crowd or fails to get in is expressed by a girl who said simply, "Money, fancy clothes, good house, new cars, et cetera: the best."

These qualities are all of a piece; they express the fact that being born into the right family is a great help to a girl in getting into the leading crowd. It is expressed differently in different schools and by different girls, sometimes as "parents having money," sometimes as "coming from the right neighborhood," sometimes as "expensive clothes." These qualities differ sharply from some of those discussed above, for they are not something a girl can *change.* Her position in the system is ascribed according to her parents' social position, and there is nothing she can do about it. If criteria like these dominate, then we would expect the system to have a very different effect on the people in it than if other criteria, which a girl or boy could hope to achieve, were the basis of social comparison—just as in the larger society a caste system has quite different effects on individuals than does a system with a great deal of mobility between social classes.

It is evident that these family-background criteria play some part in these schools, but—at least according to these girls—not the major part. (It is true, however, that the girls who are *not* in the leading crowd more often see such criteria, which are glossed over or simply not seen by girls who are in the crowd.) Furthermore, these criteria vary sharply in their importance in different schools.

Another criterion for being in the leading crowd is scholastic success. According to these girls, good grades, or "being smart" or "intelligent" has something to do with membership in the leading crowd. Not much, to be sure—it is mentioned less than 12 per cent of the time, and far less often than the attributes of personality, good looks, clothes, and the like. Nevertheless, doing well in school apparently counts for something. It is surprising that it does not count for more, because in some situations, the "stars," heroes, and objects of adulation are those who best achieve the goals of the institution. For example, in the movie industry the leading crowd is composed of those who have achieved the top roles—they are by consensus the "stars." In a graduate school, the "leading crowd" of students ordinarily consists of the bright students who excel in their studies. Not so for these high school girls. The leading crowd seems to be defined primarily in terms of *social* success: their personality, desirability as dates, clothes—and in communities where social success is tied closely to family background, their money and family.

Different Criteria for Boys

What about boys? What were their responses to this question about criteria for the leading crowd? Figure 4 also shows the boys' responses, in somewhat the same categories used for the girls. The first difference between these and the girls' responses is the overall lower frequency. The girls sometimes set down in great detail just what is required to get in the leading crowd—but the matter seems somewhat less salient to the boys.

For the boys, a somewhat different set of attributes is important for membership in the leading crowd. The responses below give some idea of the things mentioned.

> "A good athlete, pretty good looking, common sense, sense of humor."
>
> "Money, cars and the right connections and a good personality."
>
> "Be a good athlete. Have a good personality. Be in everything you can. Don't drink or smoke. Don't go with bad girls."
>
> "Athletic ability sure helps."
>
> "Prove you rebel against the police officers. Dress sharp. Go out with sharp freshman girls. Ignore senior girls."
>
> "Good in athletics; 'wheel' type; not too intelligent."

By categories of response, Figure 4 shows that a good personality is important for the boys, but less strikingly so than it is for the girls.

Being good looking, having good clothes, and having a good reputation are similarly of less importance. Good clothes in particular are less important for the boys than for the girls. Similarly, the items which have to do with parents' social position—having money, coming from the right neighborhood, and the like, are less frequently mentioned by boys.

What then are the criteria which are more important for boys than for girls? The most obvious is, as the graph indicates, athletics. Of the things that a boy can *do*, of the things he can *achieve*, athletic success seems the clearest and most direct way to gain membership in the leading crowd.

Having good grades, or doing well academically, appears to be a less sure path to the leading crowd than is athletics (and sometimes it is a path away, as the final quotation listed above suggests). It does, however, sometimes constitute a path, according to these responses. The path is apparently stronger for boys (where scholastic achievement is fifth in frequency) than for the girls (where it is eighth in frequency). This result is a little puzzling, for it is well known that girls work harder in school and get better grades than boys do. The ambivalence of the culture concerning high achievement among girls will be examined in some detail later. At this point, it is sufficient to note that such achievement is apparently less helpful to a girl for getting into the leading crowd than it is for a boy.

An item which is of considerable importance for the boys, as indicated on the bar graph, is a *car*—just having any car, according to some boys, or having a *nice* car, according to others—but whichever it is, a car appears to be of considerable importance in being part of the "inner circle" in these schools. In four of the five small-town schools—but in none of the larger schools—a car was mentioned more often than academic achievement. When this fact is coupled with the fact that the responses included not only juniors and seniors, but also freshmen and sophomores, who were too young to drive, the place of cars in these adolescent cultures looms even larger.

As a whole, how do the criteria for the leading crowd among boys differ from the criteria for girls? Several sharp differences are evident. Family background seems to matter less for boys; it is apparently considerably easier for a boy than for a girl from the wrong side of the tracks to break into the crowd. Clothes, money, and being from the right neighborhood hold a considerably higher place for the girls.

The same appears to be true for personal attributes such as personality, reputation, good looks—all of the things which define what a person *is*. In contrast, the criteria for boys include a much larger component of what a person *does*, whether in athletics or in academic matters. Such a distinction can be overdrawn, for a girl's reputation and her personality are certainly determined by what she does. Yet these are not clear-cut dimensions of achievement in which a person can actively do something; they are far less tangible. Furthermore, they are pliable in the hands of the leading crowd itself, whose members can define what constitutes a good reputation or a good personality but cannot ignore football touchdowns or scholastic honors. Numerous examples of the way the leading crowd can shape reputations were evident in these schools. For example, a girl reported:

> It is rumored that if you are in with either ——— or ——— that you've got it made. But they are both my friends. You've got to be popular, considerate, have a good reputation. One girl came this year with a rumor started about her. She was ruined in no time by ——— especially.

The girl who had been "ruined" was a top student and a leader in school activities, but neither of these things was enough to give her a place in the leading crowd. At the end of the school year she was just as far out of things as she was at the beginning, despite her achievements in school.

The matter is different for boys. There are fewer solid barriers, such as family background, and fewer criteria which can be twisted at the whim of the in-group, than there are for girls. To be sure, achievement must be in the right area—and athletics is by far the area which is more right than any other—but achievement *can* in most of these schools bring a boy into the leading crowd, which is more than it can do in many instances for girls.

Again there is the suggestion that the girls' culture is in some fashion derived from the boys: The girl's role is to sit there and look pretty, waiting for the athletic star to come for her. A girl must cultivate her looks, be vivacious and attractive, wear the right clothes, but then wait—until a football player, whose status is determined by specific achievements, comes along to choose her. This is, of course, only part of the matter, for in a community where the leading crowd largely reflects the "right families" in town, or in a school where "activities" are quite important, the girls have more independent power. Also, the fact that girls give the parties and determine who is invited puts a tool in their hands which the boys do not have.

It is as if the adolescent culture were a Coney Island mirror, throwing back a reflection of the adult society, distorted but recognizable. And just as there are variations in different places in the adult society, there are variations in different schools in the adolescent society. Their existence should be kept in mind, in order not to make the serious mistake of seeing the "adolescent culture" as all of a piece, as a single invariant entity.

Popularity with Own or Opposite Sex

In the questionnaire, boys were asked what it takes to be popular with other boys, and then, in a separate question, what it takes to be popular with girls. Girls were asked a similar pair of questions about popularity with their own sex and popularity with the opposite sex.

A comparison of the pair of questions shows an interesting result: "good grades," or doing well in school counts for something with one's *own* sex, but for very little in popularity with the opposite sex. In contrast, out-of-school activities and attributes (a boy's having a car, or a girl's having good clothes) count for much more in popularity with the opposite sex. The differences are expecially great for a girl's popularity. Doing well in school is of some value in making a girl popular with other girls, but it has little or no value in her popularity with boys.

Perhaps these results are to be expected. Yet their implications are not so obvious. Let us suppose that the girls in a school valued good grades more than the boys did. One might naïvely expect this to mean that the presence of these girls would be an influence on the boys toward a higher evaluation of studies. Yet these data say that is *not* the case; they say that a boy's popularity with girls is based less on doing well in school, more on such attributes as a car, than in his popularity with other boys. Similarly for girls—their popularity with boys is based much less on academic studies than is their popularity among members of their own sex.

The standards men and women use to judge each other have always included a large component of physical attractiveness and a smaller component of the more austere criteria they use in judging members of their own sex. Yet adults seem to ignore that this is just as true in high schools as it is in business offices, and that its cumulative effect may be to de-emphasize education in schools far more than they realize. In the normal activities of a high school, the relations between boys and girls tend to increase the importance of physical attractiveness, cars, and clothes, and to decrease the importance of achievement in school activities. Whether this *must* be true is another question; it may be that the school itself can shape these relations so that they will have a positive effect rather than a negative one on the school's goals.

The general research question is this: What kinds of interaction among boys and girls could lead each sex to evaluate the other less on grounds of physical attraction, and more on grounds which are not so superficial? It seems likely, for example, that in some private schools such as the Putney School, the kinds of common work activities of boys and girls lead to different bases for evaluating the opposite sex than do the usual activities surrounding a public high school. The question of practical policy, once such a research question has been answered, is even more difficult: What can a school do to foster the kinds of interactions and activities which lead the judgments of the opposite sex to be made on grounds which implement the school's goals?

It is commonly assumed, both by educators and by laymen, that it is "better" for boys and girls to be in school together during adolescence —if not better for their academic performance, then at least better for their social development and adjustment. But this may be not at all true —the benefits may depend wholly upon the kind of activities within which their association takes place. Coeducation in some high schools may be inimical to *both* academic achievement *and* social adjustment. Again, we should emphasize that the dichotomy often forced between "life-adjustment" and "academic emphasis" is a false one, for it forgets that most of the teenager's energy is not directed toward either of these goals. The relevant dichotomy is instead cars, clothes, and the cruel jungle of rating and dating versus school activities, whether of the academic or life-adjustment variety.

Perhaps this is where the emphasis *should* be among girls: on making themselves into desirable objects for boys. Perhaps physical beauty, nice clothes, and an enticing manner are the attributes which should be most important among adolescent girls. Yet in none of the roles in adult life that most girls will occupy are physical beauty, an enticing manner, and nice clothes as important for performing successfully as they are in high school. Even receptionists and secretaries, for whom personal attractiveness is a valuable attribute, must carry out their jobs well, or they cannot survive. Comparable performance is far less important in the status system of the high school. A girl can survive much longer on personal attractiveness, an enticing manner, and nice clothes.

The adult activities of women in which such attributes *are* most important are of a different order from those of wives, citizens, mothers, career women, secretaries: they are the activities of chorus girls, models, movie and television actresses. In all these activities, women serve as *objects of attention* for men, and even more, objects to *attract* men's attention. These attractions are quite different from the attributes of a good wife, which involve less superficial qualities. If we want our high schools to inculcate the attributes which make girls successful as objects to attract men's attention, then these values of good looks and nice clothes discussed above are just right; if not, then the values are quite inappropriate.

A second answer to the question, "What is wrong with these values?" is this: Nothing is wrong with such values, so long as they do not completely pervade the atmosphere. The values are all right, as long as there are also *other* ways a girl can become popular and successful in the eyes of her peers. And there are other ways, as indicated by the emphasis on "a nice personality" in the questions discussed above. Yet these two questions suggest that in adolescent cultures, these superficial, external attributes of clothes and good looks do pervade the atmosphere to the extent that girls come to feel that this is the only basis or the most important basis on which to excel.

52

Relationships Between Self-Acceptance and Sociometric Choices

Hayne W. Reese

In this study Reese has attempted to test the relationship between the way a child reports on himself and the way his peers rate him. It is generally assumed that such a relationship exists, but Reese's findings indicate the difficulty of scientifically validating a commonsense assumption. In a replication study, not yet published, Reese found that when he used a different population, sociometric scale, and administrative procedure, the significant correlations obtained in the present study did not hold. The importance of clarifying operational definitions for such terms as self-concept *and* self-acceptance *are indicated both in this study and in the replication.*

Several studies (e.g., Fey, 1955; McIntyre, 1952) found no significant relationship between self-acceptance and acceptance by others, but Miyamoto and Dornbusch (1956) found higher acceptance by others for subjects with high self-concepts than those with low self-concepts, and Marshall (1958) found in an experimental study that a negative shift in self-acceptance followed unfavorable reactions by the peer group, though no change followed favorable reactions. The conflicting nature of the evidence may result in part from the use of different definitions of self-acceptance. In the present study two measures of self-acceptance were used, and their relationships with several sociometric measures were investigated.

METHOD

Subjects

The subjects were 507 children from the fourth, sixth, and eighth grades of the Clarence, New York, school system. The data of 66 of these subjects were eliminated, 30 subjects missed one or more of the tests, 9 subjects obtained negative discrepancy scores, and 27 subjects named no best friend in their class. Additional subjects, as indicated below, were eliminated in a random manner to obtain proportional cell frequencies for the analyses of variance.

Tests and Procedure

The tests were Lipsitt's (1958) Self-Concept scale and his Ideal-Self Scale, a "Rate" Sociometric scale similar to one described by McCandless, Castaneda, and Palermo (1956), and a "Check-List" Sociometric scale. Lipsitt's

---

From *Journal of Abnormal and Social Psychology,* 1961, 62: 472–474.

Table 1: *Mean sociometric scores of groups based on self-concept scores*

Sociometric Measure	Group		
	L	M	H
Acceptance of Others	2.23	2.10	2.16
Acceptance by Others	2.30	2.02	2.17
Acceptance by Best Friends	1.75	1.47	1.60

Note.—Sex and grade have been collapsed since neither interacted significantly with groups. A low mean indicates high acceptance.

Self-Concept scale contains 22 trait-descriptive adjectives, each prefaced by the phrase "I am . . ." and followed by a five-point rating scale with the titles "not at all," "not very often," "some of the time," "most of the time," and "all of the time." The Ideal-Self scale contains the same adjectives, prefaced by the phrase "I would like to be . . . ," and the same rating categories.

The Rate Sociometric scale contained the names of every same-sexed member of a given class, with a five-point scale beside each name. The subjects were instructed to check 1 if the child named was a "best friend"; 2 if he was a friend, but not a best friend; 3 if the subject did not know the child very well or neither liked nor disliked him; 4 if the subject disliked him more than he liked him; and 5 if the subject did not like the child.

The Check-List Sociometric scale contained the names of every same-sexed member of a given class, and the subjects were instructed to check the names of their "best friends." Only the children who were both checked and rated 1 were counted as the subject's best friends.

The tests were given in two sessions by the teachers in their regular classes. If a child missed a session, he was tested individually if he returned to school within a week; otherwise his data were eliminated. If the discrepancy score, i.e., the ideal-self minus the self-concept score, was negative, the subject's data were eliminated (Lipsitt, 1958). The data of the subjects who had no "best friends," as defined above, were also eliminated.

For the analyses of the data, the same-sexed subjects within each grade were divided into three groups on the basis of self-concept scores, and also on the basis of discrepancy scores. Group L consisted of the lowest third, Group M the middle third, and Group H the highest third. It was assumed that the differential responses to Lipsitt's (1958) scales reflected different degrees of self-acceptance, though they may reflect differential tendencies to respond in socially favorable terms. The empirical findings are not affected by the possible social desirability component of Lipsitt's scales, though theoretical interpretation might be affected.

RESULTS

Groups Based on Self-Concept Scores

To obtain proportional cell frequencies the data of 33 subjects were eliminated, leaving 408 subjects in the groups based on self-concept scores. There were 17 boys and 17 girls in each group in the fourth and sixth grades, and 34 of each sex in each group in the eighth grade. Girls obtained a significantly higher mean self-concept score than boys (see first column of Table 2); the sex difference was larger in Group L than in Groups M and H ($t = 2.45$, 4.47), and larger in Group M than in Group H ($t = 2.02$). The sex difference was smaller in Grade 4 than in Grades 6 and 8 ($t = 2.15$, 2.61), but the latter two grades did not differ significantly in the size of the sex difference ($t < 1.00$).

The data provided two measures of acceptance of others, the mean of the ratings the subject gave others, and the percentage he named as best friends. Since the product-moment correlation between these measures was −.65 (indicating a positive relationship), and they were related to self-acceptance in essentially the same way, only the findings for acceptance of others measured by the mean rating are reported. The first line of Table 1 gives these means. The curvilinear trend was significant, as indicated in the second column of Table 2, and tests of the differences between individual means showed that Groups L and M differed significantly ($t = 2.65$), but Group H was not significantly different from L or M ($t = 1.45$, 1.20). Analysis of the simple effects in the Sex × Grade interaction showed that the sixth and eighth grade girls obtained lower means than the boys in these grades ($t = 3.56$, 2.34), but the sex difference in the fourth grade was in the opposite direction and not reliable ($t = 1.53$).

The second line of Table 1 gives the mean acceptance by same-sexed classmates, or "acceptance by others." The curvilinear trend, paralleling that for acceptance of others, was significant, and t tests showed that each group was significantly different from the other groups ($t = 4.44$

for L vs. M; 2.14 for L vs. H; 2.30 for M vs. H). The girls were more highly accepted by others than boys were in the sixth and eighth grades (t = 3.20, 2.18), but the sex difference in the fourth grade was in the opposite direction and not significant ($t < 1.00$).

The product-moment correlation between acceptance by others and acceptance by best friends was .63, and the latter measure was related to self-acceptance in essentially the same way as the former, though less strongly (see Tables 1 and 2). Groups L and M differed significantly in acceptance by best friends ($t = 3.46$), but Group H was not significantly different from Group L or M (t = 1.77, 1.69), and none of the simple effects in the Sex × Grade interaction was significant.

Acceptance by others was positively related to the two measures of acceptance of others; the product-moment correlation with the mean rating given to others was .20 ($p < .001$), and the correlation with the percentage measure was −.20 (indicating a positive relationship). Acceptance by best friends was not significantly related to the mean rating given to others ($r = -.01$), but was negatively related to the percentage measure ($r = .10, p < .05$).

Groups Based on Discrepancy Scores

To obtain proportional cell frequencies in the groups formed on the basis of discrepancy scores, the data of 45 subjects were eliminated, leaving 16 boys and 16 girls in each fourth and sixth grade group, and 34 of each sex in each eighth grade group. The sixth grade boys obtained a higher mean discrepancy score than sixth grade girls ($t = 2.55$), but the sex differences in the fourth and eighth grades were not significant (each $t < 1.00$). Group L girls obtained a higher mean discrepancy than Group L boys ($t = 2.13$), but the trend was reversed in Group H ($t = 4.66$), and the means were equal in Group M.

Analyses of the sociometric measures indicated that there was no significant Group effect; there were significant sex and grade differences, but these were not consistent between different measures.

SUMMARY AND CONCLUSIONS

Lipsitt's (1958) Self-Concept and Ideal-Self scales and two Sociometric scales were given to fourth, sixth, and eighth grade subjects.

1. Acceptance of others, acceptance by others, and acceptance by best friends were curvilinearly related to self-concept scores, with highest acceptance in a group with moderate self-concept scores and lowest in a group with low self-concept scores.
2. The trends were not significantly different in different grades or sexes.
3. Acceptance by others was more strongly related to self-concept than was acceptance by best friends.
4. The sociometric measures were not significantly related to the discrepancy between ideal-self and self-concept scores, suggesting that the discrepancy score obtained by subtraction may not be a valid measure of self-acceptance.

Table 2: *Summary of analyses of variance of groups based on self-concept scores*

Source	*df*	Self-Concept Scores		Acceptance of Others		Acceptance by Others		Acceptance by Friends	
		MS	*F*	*MS*	*F*	*MS*	*F*	*MS*	*F*
Level (L)	2	11982	704.82****	.63	3.50*	2.73	9.75****	2.62	5.95***
Sex (S)	1	1491	87.71****	1.28	7.11**	2.49	8.89***	1.05	2.39
Grade (G)	2	3	<1.00	.04	<1.00	.04	<1.00	2.18	4.95**
L x S	2	168	9.88****	.17	<1.00	.08	<1.00	.01	<1.00
L x G	4	31	1.82	.05	<1.00	.44	1.57	.94	2.14
S x G	2	62	3.65*	1.20	6.67***	.84	3.00*	1.50	3.41*
L x S x G	4	34	2.00	.40	2.22	.58	2.07	.60	1.36
Within	390	17		.18		.28		.44	

Note.—"Level" refers to groups formed on the basis of self-concept scores. The last column summarizes the analysis of acceptance by best friends.

** p < .05.*
*** p < .01.*
**** p < .005.*
***** p < .001.*

REFERENCES

FEY, W. F. Acceptance by others and its relation to acceptance of self and others: A revaluation. *J. abnorm. soc. Psychol.*, 1955, *50*, 274–276.

LIPSITT, L. P. A self-concept scale for children and its relationship to the children's form of the manifest anxiety scale. *Child Develpm.*, 1958, *29*, 463–472.

McCANDLESS, B. R., CASTANEDA, A., & PALERMO, D. S. Anxiety in children and social status. *Child Develpm.*, 1956, *27*, 385–391.

McINTYRE, C. J. Acceptance by others and its relation to acceptance of self and others. *J. abnorm. soc. Psychol.*, 1952, *47*, 624–625.

MARSHALL, R. J. Variation in self-attitudes and attitudes toward others as a function of peer group appraisals. Unpublished doctoral dissertation, University of Buffalo, 1958.

MIYAMOTO, S. F., & DORNBUSCH, S. M. A test of interactionist hypotheses of self-conception. *Amer. J. Sociol.*, 1956, *61*, 399–403.

PERSONALITY AND SELF

53

Changes in Self Concepts in Relation to Perceptions of Others

Dorothy M. Kipnis

A major concern of adolescents is self-identity. Most social psychological theories of self recognize the importance of the relationship between self-evaluation and evaluation by others. Kipnis studied a college freshman population to determine the role of person perception in the modification of self-description. This study might be examined in conjunction with those by Lamy (36), Gordon and Wood (32), and Yarrow and Campbell (37).

The self concept is a topic of investigation which is of interest from two points of view. Social psychologists are interested in the process through which individuals form and change their beliefs about themselves. Psychotherapists are interested in the relationship between the nature of an individual's belief about himself and his adjustment, and in changes in self concepts concomitant with psychotherapy.

The present investigation concerns the relationship between interpersonal perception and the process of changing self concepts. The theoretical basis for the study is developed from the theory of social comparison processes proposed by Festinger (1954). Festinger's theory deals with the formation of subjective evaluations of abilities and is based on the proposition that self-evaluations, as well as changes in self-evaluations, take place in situations where it is possible to compare one's own performance with the performance of others. The present study is an attempt to extend the theory from the area of self-appraisal of abilities to the area of self-appraisal of personality traits.

One component of Festinger's theory is that a person's evaluation of an ability that is unclearly defined (e.g., ability to write poetry) will depend to a large extent on the opinions which significant "others" have of his ability. This assumes that the individual receives "feedback" from others as to their opinions of his abilities. When we turn to the area of personality traits, however, it may be noted that there are strong restraints against discussing one's own personality directly with others. A person rarely receives direct communication from others as to what they think about his personality traits, nor is he often able to tell others what he thinks about himself. The lack of actual exchange of opinions may then

---

From *Journal of Personality*, 1961, *29* 449–465

prevent a process of social influence taking place through direct communication. The present investigation assumes that this barrier to direct discussion forces the individual to rely on his own observations of others, to compare their behavior with his own, and to use such data to make subjective comparisons as a basis for self-evaluation of personality traits.

The process of changes in self concepts will be studied in relation to individuals' perceptions of their friends.

HYPOTHESES

> 1. Individuals will perceive their friends to be more similar to themselves than others whom they like less well.

Festinger's theory begins by positing that individuals attempt to ascertain whether their abilities are good, adequate, or poor. One means of doing this is by comparing their own performance with that of other members of groups to which they belong. Festinger points out that such comparisons are not indiscriminate, but are much more easily made when the person judges that the other group members are at least fairly like himself. The greater opportunity for self-evaluation afforded by situations where others are judged to be relatively near one's own level should then mean that individuals find such situations more attractive than situations where they judge the others to be very divergent. Or, when the traits to be judged are relatively unclearly defined as are personality traits, individuals may tend to attribute similarities to others to whom they are attracted. In either case, in support of the assumption that friends serve in part as a basis for comparisons through which self-evaluations can be made, we should find that individuals tend to perceive their friends to be more like themselves than others whom they like less well.

> 2. When individuals are attracted to others whom they view as different from themselves, they show lowered stability over time in evaluations of their own personality traits.

A person may be strongly attracted to a group in which he sees all of the other members as quite unlike himself. It was previously postulated that very divergent others do not afford a good basis for comparison. If this is true, persons who are attracted to situations where they view others as very different should have difficulty in evaluating themselves with confidence. We should expect this lack of confidence to be manifested in changes or fluctuations in the person's self-appraisal over a period of time.

> 3. Individuals change self-evaluations of personality traits over time so that they are more congruent with their perceptions of persons to whom they are attracted.
>
> a. Persons who are attracted to others whose personality traits they evaluate more favorably than their own will evaluate themselves more favorably at a later date.
>
> b. Persons who are attracted to others whose personality traits they evaluate less favorably than their own will evaluate themselves less favorably at a later date.

Festinger hypothesized that the perception of differences in abilities between the self and others used as a basis for comparison should lead to attempts to reduce the differences. The present study predicts the same set of events to occur in the case of personality traits. At this point, further similarities between personality and ability traits may be noted. Festinger points out that Western society maintains values concerning abilities such that the higher a performance score, the more desirable it is. This is also true of most of the personality traits in every day usage. Characteristics such as "maturity," "friendliness," or "cooperativeness" are certainly seen as much more desirable than their opposites. Accordingly, it is expected that personality self-ratings can be ascribed along a good-bad dimension and that the direction of changes in self-ratings may be predicted along this value dimension.

> 4. Attraction to persons perceived to differ negatively from the self will be less stable over time.

Cultural standards place a premium on doing better. Consequently, individuals not only wish to assess themselves correctly—there is also definite pressure to do better and better. The value set on high performance produces a number of phenomena which are absent when it is a matter of indifference as to what kinds of beliefs one holds, so long as that belief is correct. Festinger found that individuals compared themselves predominantly with others whose abilities were superior to their own, rather than inferior, and that there were competitive strivings to excel the other person used as a basis for comparison.

If personality traits as well as abilities are differentially valued, then it can be expected that strength of attraction will differ for persons who perceive others to be superior as compared to those who perceive others to be inferior. For the person who compares himself with someone whose behavior he judges generally more positively than his own, pressure toward conformity and desire to see himself in a favorable light operate in the same direction. Thus the combined forces operating to maintain attraction in this situation are stronger than for the individual who judges another person's behavior less positively than his own. In this latter case, pressure to be like the other person will conflict with the desire to see himself favorably. These conflicting pressures are expected to produce less stable attraction over time.

METHOD

Subjects

*S*s were 87 male University of Illinois freshmen living in one of the university dormitories. The dormitory is divided into sections by partitions, most sections being occupied by four students who lived in close proximity during the semester. In addition to their roommates, most residents of the dormitory became acquainted with students who lived in adjacent sections.

Measures

Personality Description Scale. This instrument consisted of a list of 24 personality traits which were to be rated by *S*s on a seven-point scale. The items were pairs of adjectives of opposite meaning chosen to represent 12 personality factors given by Cattell (1950).

Samples of the items are as follows:

Cooperative .... Uncooperative
Mature .... Immature
Boastful .... Modest

Groups of *S*s were designated as those living in two adjacent sections of the dormitory, usually eight students. *S*s described each member of their groups, including themselves, on the Personality Description Scale.

Sociometric friendship choices. Sociometric friendship choices were made by each *S* from among members of his group. *S*s listed in order of preference the two group members they would prefer in each of 10 situations such as double dating, borrowing or lending money, as roommates, classmates, or weekend guests. A weight of two and one were assigned to first and second choices respectively. Scores were summed over all situations to give a "friendship index." A "best friend" and a "least liked roommate" were then designated for each *S*. The "best friend" of any *S* was the group member to whom he gave most choices. As no data on rejection were available, *S*s least liked roommate is the roommate to whom he gave fewest choices. The least preferred of *S*s roommates was selected in order to insure that *S*s were well acquainted with the person designated as least liked. Comparisons made by *S*s between themselves and their "best friends" were expected to constitute a salient basis for self-evaluations, while comparisons between the self and "least liked roommates" were expected either not to occur or to have little effect. Perception of differences between self and "least liked roommates" is then used for control or comparison purposes.

The Personality Description Scale and sociometric choices were obtained twice, with a six week period elapsing between testing sessions. At the time of the first session, *S*s had lived together for about five weeks.

General Statistical Procedure

D statistic. Measures of perceived difference between self and any other group member were obtained as a function of the difference between the rating on an item of the Personality Description Scale that *S* assigned himself and the rating he assigned to the other group member. The difference between an *S*'s ratings of himself at the first and second testing sessions was taken as the measure of change in his self concept. In either case, the D statistic (Cronbach & Gleser, 1953) was used as an index of the distance between the two descriptions. The index is computed by taking the difference between the two descriptions on each item, squaring it, and summing over the items. D is then the square root of this sum. . . . In terms of the index D, the lower the score, the smaller the perceived difference and hence the greater the assumed similarity.

Self-esteem. In addition to measures of differences (D), a second scoring system was used for the Personality Description Scale. As men-

Table 1: *Average perceived difference (D) between the self and friend and between the self and least-liked roommate at testing sessions I and II*

Testing session	Average perceived difference		Diff.	*N*
	Self and best friend	Self and least-liked roommate		
Time I	7.4	9.5	2.1**	87
Time II	6.5	9.0	2.5**	86[a]

** $p < .01$ by sign test.
[a] One S failed to describe his best friend at the second testing session.

tioned previously, most personality traits can be described along a desirable-undesirable dimension. In the present study three judges were asked to designate the pole of each item on the Personality Description Scale which they felt was the more desirable attribute of personality. The results of this sorting showed almost perfect agreement, one of the judges disagreeing on only one of the 24 items. A measure of self-esteem was obtained by scoring each item from 1 to 7 and summing over all 24 items. This scoring was also applied to *S*s' descriptions of others, giving a measure of esteem for friends.

RESULTS

Selection of Friends

The first hypothesis was that individuals perceive their friends to be more like themselves than others whom they like less well. This hypothesis was tested by comparing the average D score between *S*s and their best friends with the average D score between *S*s and their least liked roommate. These data for both time periods are shown in Table 1.

It may be seen that *S*s perceived less differences between themselves and their best friends than between themselves and their least liked roommates ($p < .01$ by sign test) at both Time 1 and Time 2. In short, hypothesis 1 is confirmed.

This result is in agreement with results obtained in a number of previous studies. Although neither the present study nor previous investigations allow inference of the direction of causality, they all provide strong evidence that the greater the attraction, the greater the perception of similarity.

Stability of the Self Concept

It was proposed that *S*s who were attracted to others whom they perceived to be different from themselves would have difficulty in using these others as a basis for comparison. As a result, it was expected that such persons would show less stability over time in their self-evaluations than would *S*s who were attracted to others whom they saw as like themselves. Hypothesis 2 was tested by correlating changes in self-evaluations (D scores) from Time 1 to Time 2 with differences perceived by *S*s between themselves and their best friends at the first testing session.

There is a possibility that *S*s with large changes in self concepts from Time 1 to Time 2 might differ in both initial level of self-esteem and in the general variability of their ratings, compared to *S*s with small changes in self concepts from Time 1 to Time 2. In other words, a correlation between amount of change in self concept and distance from best friend could be produced by either differences in mean level of self-esteem scores or in variability of ratings, without reference to descriptions of best friends.

To take these factors into account, two steps were taken. First, *S*s' self-esteem scores at Time 1 were partialed out of the correlation between distance from best friend and amount of change in self concept. This partial correlation allowed testing Hypothesis 2 independently of initial level of *S*s' self-descriptions. Second, a parallel partial correlation was computed between the amount of change in *S*s' self concepts and perceived differences between themselves and their least liked roommates. This partial correlation was considered necessary to test whether the first partial correlation merely reflected *S*s' consistent tendency to show variability.

These partial correlations are both presented

Table 2: *Correlations of amount of change in self concepts with perceived difference between the self and best friend and perceived differences between the self and least-liked roommate with Time 1 self-esteem (S_1) scores partialed out*

	$r \cdot S_1$	*N*
From best friend	.55**	62[a]
From least-liked roommate	.15*	53[b]

* The difference between partial correlations of .55 and .15 is significant beyond the .05 level.
** Significantly different from zero beyond the .01 level.
[a] Excludes 20 Ss who changed best friends from Time 1 to Time 2, and 5 Ss whose best friends moved out of the dormitory.
[b] Analysis based on Ss who chose the same least-liked roommate at both testing sessions.

in Table 2. It can be seen that *S*s who perceived greater differences between themselves and their best friends did tend to change their self concepts more than *S*s who perceived smaller differences. The correlation between perceived difference from the best friend and amount of change is significant at the .01 level. This correlation is also significantly higher (at the .05 level) than the correlation between amount of change and perceived difference from *S*'s least liked roommate.

The perception of difference between the self and friend is then associated with a tendency to change evaluations of the self, while the perception of difference between the self and a less liked other person shows little tendency to result in change. The results then clearly corroborate our second hypothesis.

Changes in Self-Ratings

Hypothesis 3 predicted that *S*s would change self-ratings of personality to be more congruent with their perceptions of the personality of persons to whom they were attracted. In terms of the current research, this prediction would lead us to expect a significant reduction in the mean D value between self-ratings and descriptions of best friends at Time 2 when compared to D values at Time 1.

The data to test this hypothesis are given in Table 3. In this table—columns (a) and (b)—are given the mean D values between *S*s' self-ratings and their ratings of their best friends at Time 1 and Time 2. Analyses were performed separately for *S*s who chose the same person as best friend at both time periods and for *S*s who chose a different friend at Time 2 than at Time 1.

Considering first the results for *S*s who maintained the same friendships through both time periods, it can be seen that they described themselves as more like their best friends at Time 2 than at Time 1. D values for Time 2 were significantly less than D values obtained at Time 1 ($p < .01$).

Before it can be concluded that the data support Hypothesis 3, it is necessary to show that the results were not due to *S*s describing their best friends at Time 2 as more similar to themselves—that is, that *S*s did not merely change their perceptions of their best friends while keeping their self-perceptions unaltered. Evidence on this point is provided in columns (c) and (d) of Table 3. In column (c), the mean D value between *S*s' self-ratings at Time 2 and their original descriptions of their best friends at Time 1 has been computed. This D value is significantly less than the average D value between self-ratings and descriptions of best friends, both taken at Time 1 ($t = 3.20$, $p < .01$). Thus, by Time 2, *S*s perceive themselves to be more similar to their previous perceptions of their best friends. On the other hand, it is seen in column (d), that the average D value between self-ratings at Time 1 and descriptions of best friends at Time 2 is not significantly less than the original average D value computed at Time 1. These data clearly indicate that the changes that occurred from Time 1 to Time 2 can be attributed to changes in *S*s' conceptions of themselves and not in their conceptions of their best friends' personalities.

Data for *S*s who changed friends are also presented in Table 3. While the hypotheses of the study do not lead to specific predictions for these *S*s, the analyses as applied to them gave rise to interesting and suggestive results. Two trends are apparent. First, in contrast with *S*s who maintained consistent friendships, they perceived themselves as slightly less similar to their best friends at Time 2 than at Time 1—column (b). Nevertheless, their second self-ratings have become significantly more similar to their original conception of their best friends' personalities—column (c). Thus, for these *S*s, there were changes in both their conception of themselves and their conception of their best friends' personalities. At Time 2, they described themselves as more like their original conception of their best friends, even though they no longer chose the same persons.

These *S*s can also be considered in relation to the new friends they chose at Time 2. If this is done, they perceived less difference between

Table 3: *Changes in perceived differences between the self and best friend from Time 1 to Time 2*[a]

	MEAN PERCEIVED DIFFERENCES (DISTANCE SCORES)				Comparison		
	Self and best friend		Second self-description from first description of best friend	Second description of best friend from first self-description			
	Time 1 (a) D	Time 2 (b) D	(c) D	(d) D	(a) vs. (b) *t*	(a) vs. (c) *t*	(a) vs. (d) *t*
Ss who maintained the same best friend (*N* = 61)	7.13	6.33	6.62	7.18	3.05**	3.20**	*ns*
Ss who changed best friends between Time 1 and Time 2 (*N* = 20)	8.05	8.60	7.10	9.40	*ns*	3.06**	*ns*

*** Significant beyond the .01 level.*

[a] *Five Ss chose persons who moved out of the dormitory as best friends the first session, while one S failed to describe his best friend at the second testing session. These Ss were omitted.*

Table 4: *Percentages of increases and decreases in self-esteem according to the direction of the perceived difference between the self and best friend at Time 1*[a]

	Ss with consistent friendship			*Ss* who changed best friends		
	N	Percentage increasing in self-esteem	Percentage decreasing in self-esteem	*N*	Percentage increasing in self-esteem	Percentage decreasing in self-esteem
Best friend described more favorably than self	40	70	23[b]	7	86	14
Best friend described less favorably than self	18	28	67[b]	11	45	55

[a] *Excludes five Ss who described self and best friend equally favorably, five Ss who chose persons who moved out of dormitory, and one S who failed to describe best friend at second testing session.*

[b] *The residual percentages of Ss neither increased nor decreased.*

themselves and their new best friends at Time 2 than at Time 1 ($t = 2.33$, $df = 18$, $p < .05$). This process was apparently accomplished through changes in both self-perceptions and perception of the new best friends. Neither change, however, was statistically significant when considered alone.

As a final point, it will be noted that *Ss* who changed best friends tended to perceive greater difference between themselves and their best friends at Time 1 than did *Ss* who did not change friends. The difference in D values is not significant ($t = 1.43$), but it does suggest that lack of perceived similarity might be concomitant with less stable friendships.

Changes in Self-Esteem

Hypotheses 3a and 3b predicted that *Ss* who evaluated their best friends more positively than themselves at Time 1 would describe themselves more favorably at Time 2 and similarly, *Ss* who evaluated their best friends less favorably than themselves at Time 1 would describe themselves less favorably at Time 2.

Ss were categorized according to the direction of the difference (as indicated by esteem scores) they perceived at Time 1 between themselves and their best friends. *Ss* who changed best friends were categorized in terms of perceived difference from their original best friend. Table 4 gives the percentages of *S*s in these categories who increased or decreased in self-esteem from Time 1 to Time 2.

Considering first the *S*s who chose the same best friend at Time 1 and Time 2, the results show clearly that the direction of the perceived difference between the self and best friend is related to the direction of change in self-esteem. At Time 2, 70 per cent of the *S*s who described their best friends more favorably than themselves increased in self-esteem while only 28 per cent of the *S*s who described themselves more favorably than their best friends showed similar increases. This difference is significant at the .01 level by chi-square (chi-square = 8.63, corrected for continuity, $p < .01$, $df = 1$).

The possibility existed that *S*s who described their friends more favorably than themselves were *S*s with initial low self-esteem scores, while those who described their friends less favorably were *S*s with exceptionally high self-esteem scores. Accordingly, an analysis was performed where pairs of *S*s were matched on the basis of their initial self-esteem scores. The member of each pair who gave his friend the higher rating at Time 1 was then selected, and it was tested whether these *S*s showed the greater subsequent increases in self-esteem at Time 2 in comparison to the matched *S*s who rated their friends less highly at Time 1. This analysis gave parallel results to those reported above. *S*s who described their friends more favorably showed significantly greater increases in self-esteem at Time 2 than *S*s with equal initial self-esteem scores who had given their friends less positive descriptions ($t = 2.50$, $N = 30$, $p < .02$).

The results for *S*s who chose different best friends at Time 2 than at Time 1 were in agreement with the results for those who chose the same best friends. Although the proportions were based on a small number of *S*s, and the difference was not significant, most *S*s who described their best friends less positively than themselves subsequently decreased in self-esteem, while most of the others increased.

Another way of stating Hypotheses 3a and 3b is to predict that the correlation between esteem scores of selves and friends at Time 2 will be greater than the correlation at Time 1. Table 5 shows a matrix of correlations between esteem scores for selves and best friends at Time 1 and Time 2, among *S*s who retained the same best friends. It will be seen that the correlation between selves and best friends at Time 2 ($r = .56$) is greater than at Time 1 ($r = .33$). The difference between the correlations ($.56 - .33$) is equivalent to the correlation between selves and best friends at Time 2, after removing the effects of ratings at Time 1 in each case, or $r(S_2 \cdot S_1)(F_2 \cdot F_1)$. This correlation is .29, significant at the .05 level.

That this effect was due primarily to changes in self-concepts, rather than shifts in perception of friends, is shown by comparing the correlation between selves at Time 2 with friends at Time 1 ($r = .47$) with that of selves and friends at Time 1 ($r = .33$). This difference is equivalent to the part correlation between ratings of friends at Time 1 with ratings of selves at Time 2, after removing effects of the initial self-rating, $rF_1(S_2 \cdot S_1)$. This part correlation is .33, significant at the .05 level. The analogous part correlation (comparable to the difference of $.36 - .33$ in the appropriate correlations) between ratings of selves at Time 1 and of best friends at Time 2, after removal of the effects of initial best friends ratings, $rS_1(F_2 \cdot F_1)$, is .17, which is not statistically significant.

In summary, the results of this analysis also indicate that *S*s' self concepts become more like their conceptions of their best friends.

Stability of Friendship

The last hypothesis was that less stable friendships would be formed by *S*s with persons they perceived to differ from themselves in a negative direction than where the perceived difference was positive.

Of the 47 persons who initially described their best friends more favorably than themselves, only 7 *S*s (15 per cent) changed best friends from Time 1 to Time 2. Among the 29 persons who described their best friends less favorably than themselves, 11 *S*s (38 per cent) changed best friends from Time 1 to Time 2 (see Table 4). The difference between the two proportions was significant at the .05 level (chi-square = 4.01, corrected for continuity, $df = 1$). Thus, *S*s who evaluated their best friends less positively than themselves tended to break off the relationship.

In summary, *S*s who evaluated their best friends less positively than themselves at Time 1 tended to both change best friends and to decrease in self-esteem, while *S*s who evaluated their best friends more positively both maintained the relationship and increased in self-esteem. It is interesting to consider these results in relation to the idea that individuals will be attracted to situations where they can maintain their self-esteem by contrasting themselves with

Table 5: *Product-moment intercorrelations of self-esteem and esteem for friends at Time 1 and 2*[a, b]

	S_1	F_1	F_2
Self-Esteem—Time 1 (S_1)	.77	.33	.36
Self-Esteem—Time 2 (S_2)		.47	.56
Friend Esteem—Time 1 (F_1)			.66
Friend Esteem—Time 2 (F_2)			

[a] *Matrix based upon 62 Ss who maintained the same best friends from Time 1 to Time 2.*
[b] *Correlations of .33 or better are significant beyond the .01 level.*

inferior others. At least in a population of normal *Ss*, it would seem that individuals are neither attracted to such situations, nor is their self-esteem enhanced by them when they occur. Instead, the route to more positive self-evaluations seems to be identification with a positively valued object, even though this means that the self is considered inferior to another individual.

DISCUSSION

The present investigation was based on a theory of social influence and conformity to social pressures. However, the manifestations of conformity which it predicts are somewhat different from those usually expected in investigations of social influences on self-evaluations. According to Festinger's theory, the individual's conformity is an attempt to attain or maintain comparability with others who furnish his "yardstick" for evaluating himself.

The role played by social influence in the Festinger theory can perhaps be contrasted with two other current positions. On one hand, psychologists are frequently highly suspicious of self-evaluations because they believe that social pressures mislead the individual into ascribing only socially desirable characteristics to himself. On the other, theorists such as G. H. Mead generally ignore the dimension of social desirability (1934). Instead, Mead's position is that the person accepts what he believes to be the judgments of others around him, without reference to their desirability or undesirability.

In contrast to both of these, Festinger's theory specifies that individuals are concerned with evaluating themselves favorably, but also correctly. The person's conformity is not with the opinions of others, but rather with what he perceives to be their standards of behavior, and a self-evaluation at any one point in time can furnish motivation for behavior which will allow changing the self-evaluation.

The present study has assumed that perception plays an important part in the process of conformity. If a person perceives that others' standards of behavior are higher than his own (assuming that the behavior concerned involves value judgments) the individual must either conform or continue to judge himself to be inferior. If he perceives himself to be superior, he has the task of maintaining his own level of behavior in the face of the belief that others do not maintain theirs. In terms of some of the items of the present study, he could try to be cooperative although he did not expect cooperation in return, he might try to be reliable although he did not expect his friend to be, etc. It is apparent from these examples that our theory assumed that perceived differences are based upon observations and comparisons of behavior, and that the person's perception and evaluations in turn have consequences that determine further behavior. This does not mean that the person evaluates his own behavior as an objective observer might, or that he is necessarily correct in his evaluations. It does mean that the individual's perception of others who are important to him has very real consequences for the way he behaves or attempts to behave.

One may ask why changes should occur primarily in self-perceptions rather than in perceptions of best friends. In my view this is because perceptions of friends are likely to be based upon social reality. Opinions of others are freely and frequently discussed in the absence of the person concerned. While beliefs about friends are open to continual confirmation through social interaction, the individual rarely receives similar confirmations for his views about himself and thus may be more likely to change, given discrepancies from valued others.

The relationships between *S*s' perceptions of their friends and changes in self concepts indicate strongly that interpersonal perception and self-perception are closely related. The consistency of these data on self-perception would seem to warrant further investigation of behavioral changes in a situation where more direct observation is feasible.

Changes in Adjustment in Relation to Changes in Self Concepts

A number of recent investigations of psychotherapy have been concerned with changes in self-perception as an integral part of changes in adjustment which occur during therapy. Rogers and Dymond (1954) have presented evidence that self concepts of clients changed markedly when therapy was effective. These changes in self concepts were systematically related to progress in therapy as measured by analysis of TAT responses.

The present investigation was in part designed to test whether similar changes in self concepts could be observed outside the formal therapeutic setting. The changes in self concepts noted by Rogers and Dymond were such that the client changed from very negative or unfavorable opinions about himself at the beginning of therapy toward a more and more positive self picture, so that he described himself as quite similar to the kind of person that he would ideally like to be. In the present investigation, some *S*s chose friends to whom they attributed very positive personality characteristics. These *S*s, like the clients in therapy, later attributed to themselves more positive personality traits.

The changes in self-perception in the present study were predicted on the basis of *S*s' perceptions of their best friends, who presumably were individuals with whom *S*s maintained an important interpersonal relationship. This suggests that, in the therapeutic relationship, clients' perceptions of themselves in relation to their perceptions of their therapists might be a crucial determinant of the success of therapy.

According to the hypotheses of the present investigation, one would expect that two conditions of interpersonal perception would be maximally effective in bringing about changes in self concepts of the kind noted by Rogers and Dymond. First, the client should perceive himself to be different from the therapist. If motivation to change results from the perception of difference between the self and an important other person, then a therapeutic relationship in which the client perceived his therapist to be similar to himself would be expected to result in little change.

Secondly, a client should evaluate his therapist favorably rather than unfavorably. If perceived differences between himself and his therapist are then reduced, changes will be toward a more positive self picture.

There is some evidence that both of these conditions are true of the effective therapeutic situation. In an investigation of the interpersonal perception of therapists and their clients, Fiedler and Senior (1952) obtained measures of clients' perceptions of their therapists in relation to both self-perception and descriptions of their ideal selves. Some of the therapists had been rated "good" by their colleagues while others had been rated "poor." Clients of the "good" therapists perceived greater differences between themselves and their therapists than did clients of the "poor" therapists. Clients of the "good" therapists also described their therapists to be more similar to their ideal selves than did clients of the "poor" therapists.

These results are consistent with the results which we would expect from the theory of social comparison processes. The client who does not perceive difference between himself and his therapist would be expected to show little change. Both the perception of difference and the direction of the perceived difference, on the other hand, are conditions which should be conducive to effective therapy.

SUMMARY

The purpose of the study was to investigate the effects of interpersonal perception on self-evaluations. The hypotheses tested were derived from a theory of social comparison processes which posits that self-evaluations are formulated through comparison between the self and others. If this is true, perception of the others who are used as a basis for comparison should be an important determinant of self-evaluations.

The *S*s of the study were 87 students living together in a university dormitory. Self-evaluations of their own personality traits were examined in relation to their perception of their best friends.

The results of the study were as follows:

1. *S*s perceived smaller differences between themselves and their best friends than between themselves and a least liked roommate.

2. *S*s who perceived their best friends to be relatively unlike themselves changed their self-evaluations more in a six-week time interval than did *S*s who perceived their best friends to be like themselves.

3. *S*s changed their self-evaluations during the six-week time interval so that they perceived smaller differences between themselves and their best friends. This reduction in perceived difference was accomplished through a process such that at the end of the six weeks, *S*s tended to evaluate themselves in the way that they had previously evaluated their best friends.

4. *S*s who ascribed relatively "good" personality traits to their best friends, as compared with themselves, changed their self-evaluations so that they later ascribed more positive traits to themselves. *S*s who gave their best friends relatively poor descriptions changed their self-evaluations in a negative direction.

5. *S*s who ascribed more negative traits to their best friends than to themselves broke off their friendships more frequently than did *S*s who ascribed more positive traits to their best friends than to themselves.

REFERENCES

CATTELL, R. B. *Personality: a systematic, theoretical, and factual study.* New York: McGraw-Hill, 1950.

CRONBACH, L. J., & GLESER, GOLDINE C. Similarity between persons and related problems of profile analysis. *Psychol. Bull.*, 1953, *50*, 456–473.

FESTINGER, L. A theory of social comparison process. *Hum. Relat.*, 1954. *7*, 117–140.

FIEDLER, F. E., & SENIOR, KATE. An exploratory study of unconscious feeling reactions in fifteen patient-therapist pairs. *J. abnorm. soc. Psychol.*, 1952, *47*, 446–453.

MEAD, G. H. *Mind, self, and society.* Chicago: University of Chicago Press, 1934.

ROGERS, C. R., & DYMOND, R. F. *Psychotherapy and personality change.* Chicago: University of Chicago Press, 1954.

54

Adaptive Behavior in Competent Adolescents: Coping with the Anticipation of College

Earle Silber, David A. Hamburg,
George V. Coelho, Elizabeth B. Murphey,
Morris Rosenberg, and Leonard I. Pearlin

This paper presents a literary report of research on one aspect of the transition period of adolescence conducted from a psychoanalytic frame of reference. The concept of "coping" is also used by Lois Murphy in The Widening World of Childhood *(1962), which describes the work at the Menninger Foundation. The discussion of "competence" should recall to the reader the articles by Hunt (1) and White (2).*

INTRODUCTION

The late adolescent-early adult period involves certain developmental tasks that derive from the emergence of biological sex maturity and the social requirements of adult life in a particular culture. These tasks may be met in many ways; different patterns of coping behavior may get the job done. The accomplishment of each task, and the way in which it is accomplished, may have consequences for the entire adult life of the individual and those close to him.

What are the tasks? A list of them may be long or short, depending on the level of specificity utilized. In contemporary American society —and indeed in many industrialized, urban, technologically complex cultures—certain tasks achieve prominence because they are requirements of the transition from being a child to being an adult. Adult roles, responsibilities, and opportunities are sufficiently different from those of childhood in such cultures that a major transition is involved here. During these late adolescent-early adult years, circumstances require a rather far-reaching change in one's way of life.

In the early psychoanalytic literature adolescence has been described mainly from the point of view of the resurgence of the infantile instinctual drives at this period of life. Along

---

From *Archives of General Psychiatry,* 1961, *5:* 354–365.

with the developing interest in ego-psychology, more and more attention has been focused on the ego's efforts at mastery, in relationship to both instinctual drives and external stimuli; thus Anna Freud in . . . "Adolescence" (1958) calls attention to some crucial defense mechanisms operating during this developmental period. Erikson's psychosocial theory of ego development (1959) conceives of phase-specific developmental tasks which must be mastered in each segment of the life cycle, emphasizing the mutual transactions between the individual and his society. Erikson views adolescence as a sanctioned intermediary period between childhood and adulthood, during which the lasting pattern of inner identity is achieved.

The usual source material for describing adolescence comes from studies of clinically disturbed people, either from reconstructions through adult analysis or from analyses conducted during adolescence. However, we wished to study a group of effective adolescents and to use a naturally occurring life situation as an extended laboratory through which the adolescent's adaptive behavior could also be studied. We were interested in exploring a life situation which offered an opportunity for examining adolescent behavior patterns in response to the potential challenges of new specific life tasks.

We therefore selected for study one of the commonly occurring experiences in our culture which may represent a challenge for personal growth in late adolescence: the transition from high school to the freshman year of college. If the shift from high school to college is taken as a revealing sample of the larger transition from adolescence into early adulthood, certain tasks become apparent. Prominent among these are the following: (1) separation from parents, siblings, and close friends; (2) greater autonomy in regard to making important decisions, assuming responsibility for oneself, and regulating one's own behavior; (3) establishing new friendships; (4) pressures (internal and external) toward greater intimacy and adult sexuality; (5) dealing with new intellectual challenges. Our aim was to learn how competent adolescents cope with the tasks presented by this particular life situation.

What do we mean when we say that an individual is coping with these tasks? We must decide what we are willing to call *coping behavior*. How will we recognize it when we see it? From a personality-social viewpoint, coping behavior has 2 broad components: (1) the effectiveness with which each task is accomplished; and (2) the cost to the individual of this effectiveness. Behavior may be considered to serve coping functions when it increases the likelihood (from a specified vantage point with respect to a specified time unit) that a task will be accomplished according to standards that are tolerable both to the individual and to the group in which he lives.

The time unit in this study was the 16-month period from the spring of the senior year of high school through the end of the first year of college. The project was naturalistic and exploratory in style. Relying on individual interviews with a small group of competent students who were in the process of changing from high school to college, we aimed at describing and understanding their adaptive behavior. This paper will focus on coping mechanisms these students utilized while still in high school in anticipation of the transitional experience.

CRITERIA USED IN THE SELECTION PROCESS

We recognized that there are a variety of patterns of competence and that any particular definition need not be regarded as all-inclusive. We chose to define our sample of students around certain criteria of competence which were used as the main bases for selecting them for study. We were interested in students who indicated a capacity to integrate competence in three major life areas: (1) in their academic work at school, (2) in their ability to maintain interpersonal closeness with a peer, and (3) in their ability to participate in social groups. This is not to say that this is the only way to measure or define competence, but for purposes of our study we wished to be explicit about the bases on which we did define it.

Certain steps in the selection process will clarify how these students were chosen and also define limiting factors which determine some unique characteristics of this particular sample of students. We wish to make it clear that we do not regard our sample as being in any way "representative" of the high school senior population.

These students also came from a rather similar socioeconomic background. The sample was drawn from the senior class of a high school which is located in a suburb of Washington, D.C. This suburb is a middle-class residential area, populated by a large percentage of professional people, managerial and supervisory governmental workers and a smaller percentage of less skilled workers. The per capita income level is high in comparison with the country at large.

After a talk to the senior class of the high school about the project, students could volunteer by returning a follow-up letter which was signed by themselves and their parents. From those who volunteered students were picked from the top half of the class who indicated that they were apt to be going to a college in the Eastern part of the country. This factor was introduced since we planned to visit our students at their respective colleges in a later part of the project. In the initial selection we preferred students who revealed a concentrated area of interests as reflected in their extracurricular activities. In addition, we used personality ratings of the students which were made by their teachers. For each student ratings were made by 8 teachers in regard to the following traits: motivation, industry, initiative, influence and leadership, concern for others, responsibility and emotional stability. We picked those students who had received the most favorable ratings by their teachers.

Finally, a number of students were interviewed and those who appeared overtly troubled, reported disturbing symptoms, or with whom the interviewer had difficulty in establishing comfortable rapport were excluded. Fifteen students (6 boys and 9 girls) were finally selected for a series of interviews on a weekly basis during the latter part of their senior year of high school. The selection was made chiefly on the basis of over-all impressions of the interviewer in relationship to the specified criteria of competence. This judgment was based upon the interviews plus reference to academic grades, school testing, and ratings. The findings to be presented represent a distillation of pertinent material from 6 or 7 interviews (spaced from April to September prior to the student's departure for college) with each of the 15 students selected for study.

DESCRIPTION OF THE STUDENTS IN RELATIONSHIP TO THE CRITERIA FOR COMPETENCE

These students showed satisfaction in their scholastic achievement and did not reveal any overt anxiety about their intellectual ability. Their class standing in the high school, based upon their grade records, seemed to us to reflect their intellectual endowment as measured by mental maturity tests administered by the high school. Not completely relying on these test scores, we included in the selection process an estimate by the interviewer of the extent to which he felt the student was making use of his intellectual potential.

These students showed a varied pattern of commitment to academic interests, ranging from those who stressed intellectual activities as a central value for them to those who included it as one important attribute in being "well-rounded." For the majority, effective work habits, coupled with their intellectual ability, accounted for their satisfactory academic standing.

In the area of interpersonal relationships we were interested in finding students who were competent in making and maintaining friendships with a peer, either of the same sex or the opposite sex. Our subjects were characterized by an absence of manifest anxiety in this area and by satisfaction with their close friends. We were attentive to such characteristics as the relationship having a feeling of mutual importance to both partners, a regard for the other person's welfare, the relationship not excluding other friendships, not being exploitative, and with a sense of shared interests and durability. We also looked for some experience in dating and relating to members of the opposite sex where a student indicated that his closest friend was a person of the same sex. The pattern of friendship included students with predominantly same-sex friendships, others more involved in dating, and another group who regarded their steady date as their best friend.

In their social relationships we looked for a satisfying membership in some social group, as revealed by actively participating in and feeling a part of such a group. We included in this either a companionship group or some participation in a service group at school or a group organized around a particular area of interest. These students were not necessarily the most "popular" of the senior class but they had found groups in which they could participate actively and feel accepted themselves. For a number of our students religious youth groups were important, both for their religious and social functions, with most putting greater emphasis on the latter. With about half of our students the friendship group included a circle of from 2 to 4 close friends. The remainder described a larger group of friends as constituting their predominant social group. Most of the girls in our sample belonged to more formalized social groups in the high school.

These students were very eager to cooperate in the study. In the interviews they were earnest and thoughtful in discussing their experiences. Their style of communication was characterized by a quality of frankness and minimal defensiveness. They expressed both positive and negative feelings with lively affect.

ANTICIPATORY COPING BEHAVIOR

In the latter part of the senior year of high school, the students having worked through the process of selecting and being selected by a college, were on the threshold of a new experience. In this situation, some adolescents may respond with increasing anxiety and disorganization. We had an opportunity to observe students who did not respond in this way and who were not overwhelmed by the anticipation of change. In the same way that one might seek to understand disturbance in functioning, we sought to understand the mechanisms operating in these students which maintained their psychological integrity and facilitated effective functioning in the face of challenge.

We will attempt to delineate the effective coping processes that were operating in the students. We recognize that the processes we will describe are interrelated and that the mere separating of them does an injustice to the concept of the total function of the personality. However, for purposes of description, we will consider them from different points of view, all of which are interrelated. First, we will describe general personality attributes which tend to facilitate involvement in and mastery of the new situation. Second, we will view ego operations which served to develop and maintain a self-image as a person adequate to the perceived requirements of a new situation. Third, we will consider those mechanisms which maintained distressful affective states within manageable limits. We will describe as many of the patterns as we were able to observe. This presentation, then, represents a composite of processes operating in our subjects; not all of these mechanisms were observed in all of the students.

GENERAL PERSONALITY ATTRIBUTES

The students' coping behavior will be described first in terms of general personality attributes which facilitated an active involvement in the new situation. Their behavior could be understood in terms of a motivation for actively seeking and dealing with manageable levels of stimulation and challenge in the environment.

Reaching Out for New Experience

We found the majority of our students expressed a very positive attitude toward newness. New experiences were not predominantly viewed as anxiety-laden and therefore to be avoided, but rather as desirable, exciting, and rewarding, and something to be welcomed. The tendency was not in the direction of maximal reduction of tension, but more positively in the direction of searching for stimulation and newness. The following are examples of such attitudes, from the interviews:

> I look forward to the people I'm going to meet, to the benefit I can get from talking to them and knowing new types of people.
>
> I'm looking forward to some sort of experience that's exciting.
>
> There is so much I haven't come in contact with which I want to.
>
> I'll be on my own. It's going to be a change from my normal routine. I enjoy new things. I'm not afraid of them. I'm looking forward to it.

The students saw in the new situation new opportunities and challenges which were viewed as opportunities for growth and change. For example, making new friends, being on one's own, separation from one's family and the new tasks involved in a more independent way of life were anticipated as desirable opportunities. We heard this expressed in these terms:

> It will be fun rooming with someone I don't know.
>
> I welcome the opportunity to take responsibility, and I will get plenty of chance.

Along with this desire for newness and change, students also felt that they did not wish to block themselves by remaining involved only in old relationships, as they went off to school. Thus, many of the students specifically said that they did not plan to room with high school friends or acquaintances. While they did not wish to give up these friendships, neither did they express a desire to cling to them as a way of avoiding new contacts with others. One girl, commenting on the fact that one of her closest friends is going to the same college, said:

> This will help, but I do not want to be with her all the time. I've known her for a long time and I'd rather try to make new friends. I intend to keep my old ones, but if you have someone like that you depend on too much, you don't go out of your way to meet others. I don't want to have one to depend on all the time like that.

Tendency for Activity

Another impressive characteristic of these students was their tendency to be active in facing the tasks of the transition. This reflected itself in the purposeful, highly autonomous way in which they assumed responsibilities for making preparations for going off to college. They were quite self-reliant in handling correspondence with the college, preparing their clothes, working out financial arrangements, etc. Two examples of the kinds of requirements that the students fulfilled during this summer were as follows: One of the students shopped for her clothes, did her own packing, wrote to her future roommate, brought her correspondence up to date, planned her courses after studying the catalog, and made arrangements with her parents to open a bank account in the community where she was going to college.

Another student shopped for his clothes, after his mother had helped him make a list of what he might need, corresponded with his future roommate by means of which a collaborative relationship was already begun. He attended a dinner for freshmen going to his college from this area, arranged for medical and eye examinations, and made financial arrangements with his parents. He also corresponded with one of his future advisers at the college to work out plans for earning extra money by working at the school. He reviewed the courses offered in the catalog so that he had some idea about his preferences for the first year's program.

Pleasure in Mastery

Related to this optimistic attitude about new experiences and a hunger for them, in many cases there were indications of active enjoyment of problem-solving and pleasure derived from the process of figuring things out and in mastering them. When the students talked about their work at school there were frequent references to the sense of pleasure which came from tackling problems which were reasonably difficult, somewhat beyond the students' current level of attainment, but which could be solved. It seemed clear to us that our students enjoyed tasks that were challenging provided they were not overwhelmingly so. One student commented:

> There's an enjoyment of solving difficult problems and learning new things. I feel it's also a challenge.

These attributes of reaching out for new experience, a tendency toward activity, and a pleasure in mastery have been commented on by White in his developing the concept of *effectance motivation*. White regards effectance motivation as involving ". . . satisfaction—a feeling of efficacy—in transactions in which behavior has an exploratory, varying, experimental character and produces changes in the stimulus field." He views this kind of motivation as being ". . . moderate but persistent, and in this, too, we can discern a feature that is favorable for adaptation. Strong motivation reinforces learning in a narrow sphere, whereas moderate motivation is more conducive to an exploratory and experimental attitude which leads to competent interactions in general, without reference to an immediate pressing need" (1959).* We found this concept useful in understanding a quality in the personalities of our subjects. We viewed effectance motivation as providing a general press toward active involvement in dealing with the new situation. We will now consider other more specific mechanisms operating in our subjects.

DEVELOPING A SELF-IMAGE AS ADEQUATE TO THE NEW SITUATION

These students viewed themselves in a positive way. However, having an image of oneself as an adequate person is not a psychological state that exists outside of the pressing challenges of life experience. We view it rather as a dynamic process related to the requirements of external reality. During the period of anticipation of college we observed certain processes that served to develop and maintain a self-image as a person adequate to master the perceived requirements of the new situation. This represents a process of matching what one perceives as being required in the new situation with an image of oneself as being competent to deal with the new situation effectively. We observed a variety of behavior patterns that facilitated the development of such a self-image. These patterns will be described next.

---

* This article appears in this volume—*Editor's note*

By Referring to Analogous Past Experiences

In maintaining their sense of competency as to their readiness for the next step, our students often referred to relevant past experiences of their own which had been adequately mastered. In effect, the students seemed to identify elements in the new situation with some experience in their own past which they felt was analogous. By reference to their ability to handle this analogous experience in the past they could reassure themselves about their ability to handle these situations in the future.

These students, in talking about the change to college, often made reference to their experience in shifting from junior high school to high school. They recalled their ability to cope with this previous transition and made comparisons with the present situation as having features in common.

One girl, who has always had her own bedroom, reminded herself that when a girl friend came and stayed with her for a month, it worked out well and therefore she felt confident that she would be able to get along with her roommate at college. Another student reminded herself that she had had experiences with moving that had helped her to learn how to make friends and taught her to deal with strangers. These references presented for the student evidence of his ability to deal with similar situations in the past and thus provided a basis for confidence in dealing with them again in the future.

By Referring to Continuity with the Present Self-Image

These students had had a series of experiences which they regarded as relevant in preparing them for the situation at hand. They also had an image of themselves as young people ready for a new situation and able to meet the challenges ahead. They expressed in various ways the feeling that collge represented a gradual part of a continuous process of growing and maturation. They tended to see a narrowing gap between their self-image and their image of college students. Long in advance of the actual departure for college a pattern of *anticipatory detachment* from parents supported the students' self-images as people ready for change. These students had a basis for confidence in their ability to live more in a world of their peers at college, since they already had had a wide experience in making friends and in establishing and maintaining meaningful relationships outside of their families while they were still in high school.

Our students expressed a feeling of readiness to grow away from their families and described a feeling of gradualness in this regard. They already had had previous experiences in being separated from their families and saw themselves in high school as young people moving in the direction of being more autonomous in relation to their parents. They felt ready to enter the next phase of their life experience. Thus change feels continuous with their present self-image. One student expressed this succinctly by saying:

> Going away is just another step on the ladder; it is not like jumping off a ladder.

For some students this anticipatory detachment extended to some of their existing ties with peers at high school. Thus, these students felt ready both for further separation from parents and for a shift in their peer relationships as well. We recognized that this preparatory behavior may have also served the function of warding off feelings of depression associated with loss.

By Learning About the New Situation in Advance

Another part of matching the self-image with the requirements of the new situation is to find out more in advance about what will be required in the new situation and what it will be like. Confidence tends to grow as the ambiguity of the new situation is reduced. Our students used a variety of channels for getting advance information about the new situation. They learned from correspondence with the college, through catalogs, talks with college friends, visiting college campuses, reading books about college, through counselors, teachers, friends, parents, and college representatives.

With those students having older siblings who had gone off to college, we heard references to learning about the experience through them and thus adding to a sense of preparation for the experience itself. For example, if there was some concern about how their parents would deal with the separation the experience of an older sibling leaving for college was brought to mind. One student recalled how her parents had been able to adjust to her brother's leaving for college. Siblings and friends served as important sources of information about college. Whether or not the ideas that the students gleaned were accurate, at least having this information served to reinforce a feeling that they were prepared for the new environment.

By Role Rehearsal

As College Students. Most of our students seemed to be preparing themselves for college by rehearsing forms of behavior which they associated with college students. We could understand this as a process of attempting to deal with part of the experience in advance by rehearsing in the present the image of a "college student." This process is related to the pattern of *anticipatory socialization* as described by Merton and Kitt: "For the individual who adopts the values of a group to which he aspires but does not belong, this orientation may serve the twin functions of aiding his rise into that group and of easing his adjustment after he has become part of it" (1950).

We saw numerous instances of this pattern of role rehearsal. Quite a number of the students reported that they began to read books that they felt were either required in college or would be useful for them in their college courses. A number of boys began buying their own clothes for the first time, saying they knew they would have to do this in college. One student who felt that scheduling would be an important part of the college experience tried to prepare herself for this by teaching herself to get things done in a minimum of time and not procrastinating or prolonging things.

Another boy, who viewed being collegiate in terms of becoming a gentleman, tried to act as he thought a college student would. He described to the interviewer his efforts at doing assignments, carrying out his responsibilities, and being a gentleman. For him, this also including letting adults through the door first, speaking when spoken to, and being more polite with girls.

Some of the special courses offered in high school were viewed as trial college experiences, thus simultaneously providing intellectual skills, bolstering the student's self-image, and coping with anxieties about intellectual abilities in this area. Students made references to these courses as "being on a college level." Some students referred to the experience of writing term papers as being an excellent preparation for college. For the most part they seemed eager for the kinds of experiences in high school which they felt would be relevant to what would be expected of them academically at college.

One of the students described in detail his experience in writing a term paper which he felt was like what he would be doing in college. He had the feeling that because of this experience and particularly because his teachers had told him that this was what students would be doing at college, he had a head start.

As Adults: In addition to rehearsing college student behavior, there was also some rehearsing of functioning more like an adult before going off to college. This seemed to be related to getting more experience in mastering tasks in the adult world and strengthening the feeling of confidence to deal with adults on a more nearly equal level. We were very much impressed with the preparatory value of the summer job for these students. Those who worked felt that this experience helped them to develop confidence in dealing with adults. This went along with a shift in their own self-image with respect to the new adult role possibilities and helped to confirm their ability to handle new situations. For many students, this was their first experience in competing for jobs in the adult world. They were no longer working at jobs associated for them with adolescent status such as baby sitting, lawn mowing, etc. Many of them took competitive examinations for summer government positions. One girl said of her government job:

> When you walk through that door you're on your own. If you don't make good, you're fired. It's not like school, where you're answerable to your parents through your teachers. I enjoy the experience of earning my own money and being on my own.

Another commented:

> We were all equals, we were all working side by side. I was glad they called me by my last name. Otherwise it would have been like being a baby-sitter. It's the first time I've been almost looked on as a contemporary.

A student who planned to enter the ministry felt that his summer job was a broadening experience for him socially and would be an asset to him when he became a minister. He had previously worked at a gas station, but in the summer before going to college obtained a government job. He commented:

> This gives me an opportunity to use my mind and associate with a different class. As a minister I will have to learn to associate with all kinds of people.

Thus, his job experience was regarded as a rehearsal opportunity in terms of future vocational goals as well.

Another student pointed out that in addition to earning money, she has learned that she is capable of meeting people and getting along with them easily. She enjoyed the feeling of responsibility. It was a novel and gratifying experience to have people give her jobs to do and leave her to her work with the understanding that she is capable of going ahead on her own.

By Group Identification

Another mechanism for the students' self-assurance that they would do well at college was to identify themselves as part of a group which shares a reputation for being adequately prepared for college. This was a view that our students had of their high school. These students, it is true, were drawn from a group of the more competent in academic performance and viewed themselves as being adequately prepared for college in terms of their own intellectual skills. In addition, however, they recognized they came from a high school which sent a high percentage of students to college. They expressed a feeling of pride in their academic preparation and revealed a feeling of confidence that attending this school had equipped them to deal adequately with the demands ahead. Many of them expressed the feeling that coming from their particular high school they were a jump ahead of students coming from most other high schools.

By Lowering the Level of Aspiration

Another mechanism had to do with students lowering their sights in terms of what they were going to expect of themselves in college. It was as if they were redefining an acceptable self-image that would permit them to maintain a feeling of satisfaction in their performance if it were at a lower level than in high school. This was most often expressed around concern about grades in college. This concern was fairly widespread and some of the students coped with it by lowering their expectations of their performance at college. This provided for them a cushion against possible future disappointment.

One student expressed rather well what several others felt, when she said that she wasn't really making firm plans about anything in college because

> If you set up certain standards for yourself, for instance belonging to a certain group, or making certain grades, or anything else, you're apt to be disappointed. It's better to be prepared to accept whatever comes your way.

We feel that this attitude did not reflect an abject resignation to the anticipated higher demands of college. Rather, these students conveyed to us that they would continue to try to maintain their academic standing. In the context of recognizing higher standards and the presence of other good students at college they prepared themselves for the possibility of a lower level of academic status.

By Selectively Perceiving Encouraging Elements in the New Situation

Our students tended to perceive college as a potentially friendly environment. This view of the unknown situation could facilitate a feeling that one would be adequate to mastery of the situation. Certain favorable characteristics were attributed to the college. Colleges were viewed as benign and wise institutions and particularly as places where support would be offered if needed. One student expressed the feeling that college would not be a sink-or-swim proposition, since he was sure he could always depend on the college advisers. He had the feeling that there would always be people to turn to if necessary.

Another attitude that was expressed was that college would be effective in placing students with compatible roommates. Some of the students spoke of their feeling that the college made efforts to place students with people who had similar interests. This view of the new situation helped sustain a feeling that there would be forces at work in the new environment which would facilitate one's being able to get along well.

MAINTAINING DISTRESS WITHIN MANAGEABLE LIMITS

In the preceding section we have described a variety of ways in which students reassured themselves that they were the kind of a person who would be able to manage the college experience well. It was not true, however, that our students approached this new experience without feelings of distress. We were impressed, however, that they were able to experience anxiety and contain it within manageable limits, so that it never became overwhelming to them. This capacity to contain anxiety has important consequences. A student who is able to experience anxiety without becoming overwhelmed by it is in a position to function more effectively and thus enhance the likelihood of using his assets in a more constructive way. For this reason we were interested in looking at those mechanisms which the students used in keeping anxiety within manageable limits.

The Supportive Function of Shared Experiences

Almost all of the students expressed an attitude that what they were experiencing in the way of anxiety was not unusual; rather, it was viewed as something that the majority of their friends in the same situation were also feeling. Thus, being aware of anxiety did not make the student feel uncomfortably-different from others. Our students' close relationship with their peers made it possible for them to communicate their feelings to others and provided a situation in which they could be in touch with what others were feeling. In this way an awareness of anxiety did not trigger additional concerns about "What's wrong with me that I'm feeling this way?" The view of "We're all in the same boat" was one of the most common means of reducing anxiety. One student, in talking about the feelings of apprehension he felt just prior to leaving for college, said that he could dispell his apprehension with the thought:

> Everybody feels this. Everybody has to be a freshman once. Everybody's in the same boat that I am and there's nothing to really worry about.

In regard to concerns about making friends, these students felt that other freshmen at college would be just as concerned as they. They recognized this as a situation which would therefore be conducive to making friends. That is, others would have as much need for making friends as they themselves felt. This was illustrated in the following quotations:

> I feel confident in my ability to make friends. In college everyone's in the same boat. There's no reason to feel I won't make friends.
>
> There we'll all be in the same boat. We'll all be out to make friends.
>
> Everyone is in the same boat. I don't think making friends will be that difficult. Everyone will want to get to know people.

The Usefulness of Worry

. . . Worrying can be viewed as adaptive when it contains a preparatory recognition of potentially difficult aspects of a life situation along with an opportunity to rehearse ways of dealing with the new situation. The students in our study did exhibit varying degrees of concern about the situation ahead; however, these concerns did not lead to a circular ruminative preoccupation.

Some of the students expressed an attitude that worrying about things had a useful function. For one student, certain magical implications were attached to worrying which then made the concern feel less threatening. That is, worrying about something somehow insured that things would work out well, as if one thereby extended his control over the uncertainties in the unknown situation. This seemed to be a way of allowing oneself to experience the distress without being overwhelmed by it. This quote from one of the interviewers illustrates this process:

> If you don't worry about something, then it will go wrong. As long as you worry a little about it, it will work out. . . . If you are overconfident and don't worry about some things, then I feel something's bound to go wrong, so that worries don't tend to bother me so much.

Present Activity in Anticipating Future Concerns

Some of our students' behavior seemed important, not only in dealing with elements in the present situation, but also in dealing with an anticipated future concern. For example, one student had some concern about whether his widowed mother could handle the household chores in his absence. He tried to get everything done around the house, including redecorating the basement and a number of little jobs that needed tending to before he left for college. Such behavior could be viewed as having future value in the new situation. It provided a reference point for handling anticipated guilt he might experience about leaving his mother while at college.

Two of our students, who had a close enough relationship during high school to consider marriage, made a decision to go to separate colleges; to date other people and "see what happens." As the girl expressed it, she felt even if they were to get married, if they hadn't had this experience, in the future she might have the thought "think of what would have happened if we had not been separated." She felt this would be helpful to their future relationship. To postpone a definite decision for a prescribed period of time provided a moratorium on making this decision immediately in college and thus free her to deal with the new situation more actively. It also was a way of dealing with an anticipated future concern.

The value of the summer job experience has already been mentioned in an earlier section of this report. We felt that another useful aspect of the job experience was that it could provide some reference point in the future when there might be some concern about the element of dependency in his role as a student. That is, another value of the work experience might be to provide a proof of ability to function in a more independent and self-sufficient way before continuing in the role of a student.

Rehearsal in Fantasy of Future Behavior

Another way of dealing with some anxieties about the new situation was to rehearse how one might deal with some future contingency by imagining ahead of time what kind of behavior could deal with it. For example, one student had enjoyed her activity in a madrigal group in high school which was an important source of satisfaction for her. In discussing her anticipation of college, she spoke of hoping to find such a group in the new setting. In exploring this in fantasy, she said that if there were no such group available, that she could picture herself doing something about organizing one.

A more common area in which this mechanism was in operation was in regard to concerns about separation from home. One student imagined herself possibly needing help at college and she fantasied that she could turn to relatives not too far from the college she had chosen. Another student saw herself being more active about becoming more involved with her peers at college. She spoke of this in the context of her feelings that she will miss her family while at college and in the interview reassured herself by saying:

> I suppose I'll become more familiar with people my own age.

Other students recognized that they would not actually be too far away from home and reassured themselves about any anticipated homesickness with the thought:

> If I want to go home. I'll be able to.

SUMMARY

As an opportunity for exploring certain aspects of ego functioning in late adolescence, we selected for study a group of competent high school seniors who were anticipating the transition to college. During the spring and summer months preceding their departure for college, we observed a variety of coping behavior patterns in these students being mobilized in regard to the tasks ahead. These many processes served various ends simultaneously and tended to overlap and reinforce each other.

A general characteristic of our students was their tendency to reach out for new experience, a tendency to be active in dealing with challenge and an enjoyment in the sense of mastery. This pleasure in effectance served to facilitate an active involvement and pleasure in coping with the new situation.

In addition, more specific mechanisms were observed. Through a variety of ways, these students developed and maintained an image of themselves as adequate to the perceived requirements of the new situation. Students referred back to analogous situations in the past which had been adequately mastered, thus reassuring themselves about their ability to handle these situations in the future. They referred also to their present self-image as evolving gradually in desired directions, and through anticipatory detachment saw themselves as ready to establish new relationships at college. By seeking out information about the new situation, they reduced some of the ambiguity in it and, in so doing, felt better prepared to deal with it.

By role rehearsal they prepared for the new situation by rehearsing in advance forms of behavior which they associated with college students. In addition, they rehearsed behaving more like adults before going off to college. This summer job experience confirmed their self-image as people adequate to master tasks in the adult world. It also strengthened their feelings that they could deal with adults on a more equal level. These students viewed themselves as members of a group which shares a reputation for being adequately prepared for college. They also dealt with concerns about performance at college by redefining an acceptable self-image that would permit them to maintain a feeling of satisfaction in their performance if it were at a somewhat lower level than in high school. In addition, they tended to perceive selectively encouraging elements in the new situation; by regarding college as a potentially friendly environment, they were sustained in feeling that one would be able to get along well there.

Other mechanisms seemed understood best as ways in which distress was contained within manageable limits. Our students were not without anxiety about college, but derived considerable support from an awareness that others were experiencing anxiety as well. "We're all in the same boat" was an attitude that was most common in maintaining anxiety from becoming overwhelming. Experiencing anxiety did not in turn signal a feeling of uncomfortable distance from others, but rather as something that was shared. Some students viewed their worrying as something useful, as if worrying extended control over the uncertainties in the unknown situation. Some of the students' behavior before going to college could be understood as providing a reference point for dealing with anticipated distress in the future. The knowledge that one was capable of earning money and holding down a job in the adult world could be a useful reference point in dealing with future concerns about the element of dependency in the role of student. Anxiety about some future contingencies could be dealt with also by rehearsing in fantasy how one would deal with that situation ahead of time.

In conclusion, during the period of anticipation of the new college experience, a variety of patterns are in process which serve to reinforce one another in a useful way. Attitudes of pleasure in dealing with newness, a process of matching the self-image with what is anticipated ahead, and mechanisms for containing distress reinforce one another to provide a general basis for confidence in facing the new situation. The active search for manageable levels of challenge in newness is more characteristic of the coping behavior of competent adolescents than a stabilized adaptation to the environment with maximal reduction of tension.

REFERENCES

ERIKSON, E. Identity and the life cycle. *Selected Papers, Psychol. Issue.* New York: International Universities Press, 1959. Vol. 1, No. 1, pp. 101–164.

FREUD, ANNA. Adolescence. *Psychoanal. Study Child*, 1958, *13*, 255–278.

MERTON, R. K., & KITT, A. S. Contributions to the theory of reference group behavior. In R. K. Merton & P. F. Lazarsfeld (Eds.), *Continuities in social research.* Chicago: The Free Press of Glencoe, 1950. P. 87.

WHITE, R. W. Motivation reconsidered: The concept of competence. *Psychol. Rev.*, 1959, *66*, 297–333.

55

Consistency and Wisdom of Vocational Preference as Indices of Vocational Maturity in the Ninth Grade

Donald E. Super

Super's study confines itself to one aspect of the adolescent's total personality—vocational preference. This particular research indicates that vocational preferences expressed by ninth-grade boys are not realistic, consistent, or wise. Super has developed a theory of vocational choice that ties in to general personality development, and in the conclusion of this paper, he suggests how vocational test data may be utilized in self-development, in spite of its lack of validity for indicating ultimate vocational preference.

The practice of inquiring concerning the vocational preferences and ambitions of junior and

---

From *Journal of Educational Psychology,* 1961, *52,* 35–43

senior high school students in helping them to make educational or prevocational choices and plans is widespread and based on compelling arguments: it would be unthinkable, in a democratic society, to require the determination of the directional choices which students make in selecting college preparatory, commercial, trade, home economics, agriculture, and general courses, without taking into account the goal to which the youth aspires. High school guidance questionnaires to be filled out by entering pupils ask what occupation they hope to enter, and numerous studies have been made of ways in which to elicit information concerning vocational preferences.

The consistency, and particularly the wisdom or realism, of vocational preferences have often been used as measures of the effectiveness of vocational guidance programs by practicing counselors and by educational and psychological research workers. Having a vocational objective is important in a society in which earning a living is important, in which occupational roles are of major significance, and in which education is, in effect if not avowedly, occupationally oriented: having a vocational preference, in this context, gives purpose to behavior and makes possible educational and vocational decisions. It has been argued that *consistency of vocational preferences* shows intensity and validity of interest, and that it is better to work consistently toward one clear-cut goal than wastefully to keep shifting objectives. In the case of *wisdom of vocational preferences*, the reasoning is that realistic goals are by definition attainable, whereas unrealistic or unwise goals are by definition those which one is not likely to attain or with which one is not likely to be satisfied if one does attain them.

Studies using consistency of vocational preferences are not very common, this being typically a counselor's method of judging how seriously to take a student's expressed preference, but the method is illustrated by Rothney's recent Wisconsin Guidance Study (1958).

Wisdom or realism of vocational preferences has been used as a criterion of the effectiveness of vocational guidance programs in many studies. Indices of realism are often used by counselors in judging students' and clients' need for guidance.

Despite the widespread acceptance of the importance of consistency and wisdom of vocational preferences, many writers on vocational guidance and on vocational development have questioned the significance of expressed vocational preferences in early adolescence. Fryer's (1931) review pointed clearly to the conclusion, confirmed by Carter's later review (1944), that the expressed preferences of boys and girls in their early and middle teens are unstable. More recently still, Schmidt and Rothney (1955) reported convincing evidence on the instability of expressed vocational preferences from one year of high school to the next and into the first year out of school: only 49% of the "choices" of the tenth grade remained the same in eleventh grade, and this figure was reduced to 35% in the twelfth grade, and to 24% in the year following graduation. . . .

In view of the persistent use of the concepts of consistency and wisdom of vocational preferences despite their tendency to be unstable in early adolescence, it is important to examine their psychological significance. In the Career Pattern Study of the Horace Mann–Lincoln Institute of School Experimentation several such measures were developed (Super & Overstreet, 1960) and related to other variables in the ninth grade.

METHOD

Subjects

The subjects of this study were the core group of the Career Pattern Study of the Horace Mann–Lincoln Institute of School Experimentation, 105 ninth grade boys who were found to be typical in age, intelligence, socioeconomic status, and other key variables of ninth grade boys in Middletown, New York, in the early 1950's. As Middletown itself is an average town on a variety of social and economic indices (Super & Overstreet, 1960), these boys may be considered typical of ninth graders in many American communities. All of the CPS boys indicated, in their interviews, at least one tentative vocational preference.

Measures of Consistency and Wisdom of Preferences

Three measures of consistency of vocational preferences were developed, the third measure being essentially a combination of the first two. The expressions of vocational preferences were obtained in tape recorded interviews in which the subjects of the study were asked " . . . about your plans for the future. What would you like to be by the time you're thirty?" This question was followed up with nondirective leads designed to keep the boy talking about his vocational preferences, and with probes designed to make sure that relevant information was obtained if not offered spontaneously (Super, Crites, Hummel, Moser, Overstreet, & Warnath, 1957). In the analysis of the data the first four vocational preferences expressed (if that many were forthcoming) were tabulated. (It should be noted that these are measures of consistency of first and alternative preferences at one point in time, not consistency of first preferences over time.)

An index of *Consistency within Fields* was obtained by classifying the expressed preferences according to the occupational field classification developed by Roe and modified by Moser, Dubin, and Shelsky (1956). The total number of fields into which the boy's preferences fell, minus 1, was his index of Consistency within Fields.

A second index, *Consistency within Levels,* was developed by classifying the same preferences according to the modified Roe occupational level scale, the score being the total number of levels at which expressed preferences fell, minus 1.

A third index, *Consistency within Families,* was developed, by summing the scores on the above two indices, an occupational family being defined as a combination of field and level (e.g., technical occupations at the professional level constitute a family by this definition).

Four measures of wisdom or realism of vocational preferences are relevant to this study. Although no one of these can be viewed as a sufficient index of wisdom of choice, each of them involves a variable which is widely accepted and objectively justifiable as one measure or criterion of realism.

The first index of wisdom or realism was one of *Agreement between Ability and Preference,* or, more accurately, of agreement between the measured intelligence of the individual and the intelligence characteristic of persons employed in the occupation of his first preference. If the boy's score on the Otis Quick-Scoring Mental Ability Test, converted into an AGCT equivalent, exceeded that of the bottom quarter of the men in his preferred occupation as shown by the manual for the Army General Classification Test, his ability was considered to be in agreement with the occupational intelligence requirements; if it was equal to that of men in the bottom quarter, his ability was considered not to be in accord with the requirements.

The second wisdom measure was the index of *Agreement between Measured Interests and Preference.* Interests were measured by the Strong Vocational Interest Blank, slightly modified to insure comprehension at the ninth grade level, and each boy's *interests* were classified as primary, secondary, or tertiary in the family in which his first *preference* fell, using Darley's (1941) method of classifying interest score patterns. When the measured interest pattern was primary in the field corresponding to the expressed preference, a score of 4 was assigned, when the pattern was secondary the boy was given a score of 3, etc.

The third wisdom index was that of *Agreement between Occupational Level of Measured Interests and Level of Preference.* The socioeconomic level of the boy's interests (occupational interest level) was measured by the Strong OL scale. If the boy's OL score was not more than one standard deviation below the mean OL score of his preferred occupation, as shown on Table 50 in Strong's monograph (1943, p. 192), the interest and preference levels were considered to be in agreement.

The final wisdom measure to be discussed here is that of the *Socioeconomic Accessibility of Preference.* The family bread-winner's occupation was rated according to the Hamburger revision (1958) of the occupational rating scale included in the Index of Status Characteristics (Warner, Meeker, & Eells, 1949). The boy's vocational preference was rated on the same scale, the agreement or disagreement of the two ratings was ascertained, and the size of the discrepancy was the index of socioeconomic accessibility.

RESULTS

It should be emphasized that no one of these indices is considered completely satisfactory, but that there are good arguments in support of each of them. For example, the occupational intelligence data of World War II are recognized as an imperfect sample of civilian occupations, but they are the best available and have proved generally usable; furthermore, the use of the first quartile as the cutting point takes into account the great range of intelligence which makes possible success in any one occupation. In the case of socioeconomic accessibility, it may be objected that it is unwise, in a fluid and democratic society, to judge the realism of a vocational preference by its correspondence with the status of the parents; but even in our democratic society many studies show that the occupation entered by the child tends to be at the same socioeconomic level as that of the parent (Super, 1957), and even when the child changes socioeconomic levels he does it with the help or hindrance of

Table 1: *Intercorrelations of indices of consistency and wisdom of ninth-grade vocational preferences*

Index	Consistency			Wisdom		
	Field	Level	Family	Aptitude	Interest	Level
Consistency:						
Within Fields						
Within Levels	22*					
Within Families	62[a]	81[a]				
Wisdom in terms of:						
Aptitudes	16*	06	18*			
Measured Interests	—01	—12	—12	—06		
Measured Interest Level	26**	04	14	00	—32	
Accessibility	08	13	18*	27**	—09	17*

Note.—N = 105 boys. Decimal points have been omitted.
[a] *Spuriously high because of use of one measure in the other.*
* *Significant at the .05 level, one-tailed test.*
** *Significant at the .01 level, one-tailed test.*

his parents' resources, contacts, information, and values. Each of these measures of the wisdom of a vocational preference is, therefore, one of the possible components of a conceptually more satisfying and valid global index of realism, in which having the appropriate intelligence might, for example, offset not having the socioeconomic background which would be a help in achieving one's ambitions.

Agreement among Indices of the Consistency and Wisdom of Ninth Grade Vocational Preferences

If consistency and wisdom of vocational preferences are valid concepts with which to work in dealing with ninth graders, one should find at least a moderate degree of agreement (correlations of .30 to .50) between various measures of consistency of vocational preferences, a similar degree of agreement among indices of the wisdom of the vocational preference, and some agreement (correlations of .20 to .35) between consistency and wisdom indices. Table 1 reports the intercorrelations of seven CPS indices of consistency and wisdom of vocational preferences, for the 105 ninth grade boys in the core group.

The indices of Consistency of Field and of Consistency of Level are, of course, highly correlated with the index of Consistency of Families: this happens because the last-named index is a combination of the first two. Only Field and Level in this group of three measures are so constructed as to be operationally independent of each other, so that the correlation between these two measures has meaning; it is .22, significant at more than the .05 level but below the .01. This suggests a slight tendency for the ninth grade boys who are most consistent in aspiring to occupations which are at the same level to be the most consistent in aspiring to occupations which are in the same general field. The implication of this very slight relationship is that the concept of "consistency of vocational preferences" has minimal meaning when boys are in the ninth grade, although it may take on more significance at a later age.

The Consistency indices, Table 1 reveals, are not at all related to Wisdom as measured by the index of Agreement between Interests and Preferences, and generally unrelated to the other Wisdom indices. Consistency of Field and Consistency of Family (based partly on Field) are related, to a low and just barely significant degree, to Agreement of Aptitudes and Preference, and Consistency of Family is similarly related to Accessibility of Preference. The correlation of .26 between Consistency of Field and Agreement between Occupational Level of Measured Interests and Level of Preference is presumably an artifact, as each of these indices is in part based on other measures which deal largely with level: Field involves type of job, and some types of jobs are largely at lower levels (e.g., Business Contact, Outdoor) whereas other types are largely at higher levels (e.g., General Cultural, Science), and Level of Interests or Preferences is by definition and operationally a measure of socioeconomic level. The relationships between Consistency and Wisdom shown by Table 1 may therefore be viewed as negligible.

The internal agreement among the indices of Wisdom of Vocational Preference must next be considered. Agreement between Aptitude and

Preference is unrelated to Agreement between Measured Interest and Preference, or to Agreement between Level of Measured Interests and Level of Preference; its correlation of .27 with Accessibility of Preference should probably be viewed as an artifact in view of the facts that the nonpreference variables in the two Agreement measures (intelligence and socioeconomic status) correlate .27 in this sample, and the preference variables (socioeconomic level of preference and intelligence level of preference) in the two measures presumably have an even higher intercorrelation because in each case one variable (preference) is scaled on one of two other highly intercorrelated variables (socioeconomic status and intelligence).

Agreement between Measured Interests and Preference is unrelated to the other Wisdom indices, two of the intercorrelations being nonsignificant and the third (−.32) being the opposite direction from that hypothesized. And, finally, Agreement between Level of Measured Interests and Level of Preference is also to be viewed as unrelated to the other Wisdom indices, as the barely significant and very low correlation of .17 with Accessibility can be attributed to an artifact: both components of both agreement measures are scaled as to socioeconomic level, and the socioeconomic level of the vocational preference is one of the two measures entering into each of the indices.

The only conclusion that it seems legitimate to draw from Table 1 is, then, that the few seemingly significant intercorrelations may be the products of artifacts, and that the various indices of consistency and wisdom of vocational preferences are unrelated to each other. This lack of relationships suggests a lack of validity as indices of anything significant in the vocational development of ninth grade boys.

Agreement Between Consistency and Wisdom of Preferences and Other Variables

The construct validity of a set of measures of the same variable depends not only upon the intercorrelations of these presumably similar measures, but also upon their agreement with other variables to which theory would lead one to expect them to be related. If consistency and wisdom of vocational preferences are considered to be indices of vocational maturity, of the degree of vocational development which has taken place in an adolescent, then they should be related to measures of other characteristics which might be expected to develop concomitantly with, to result in, or to be the product of, vocational maturity.

The Career Pattern Study therefore used or developed a series of measures of other variables which could be hypothesized as related to vocational maturity. These were age (except that the fact that the negative correlation between age and intelligence in any one grade—in which the older pupils tend to be the retarded and the younger tend to be the accelerated—and the limited age range may be expected to confuse the relationship), socioeconomic level, intelligence, school achievement, patterning of vocational interests, emotional adjustment, and peer acceptance.

Socioeconomic level was measured by the placement of the family bread-winner's occupation on the Hamburger revision of the Warner scale, intelligence by the Otis Quick-Scoring Test of Mental Ability, school achievement by grades in the three constant courses (taken by all students) of the ninth grade, patterning of interests by the application of Darley's (1941) method to scores on Strong's Vocational Interest Blank, adjustment by Overstreet's method of deriving a total adjustment score from stories told in the Thematic Apperception Test (Super & Overstreet, 1960), and peer acceptance by a modification of the Guess Who technique (Super & Overstreet, 1960).

The correlations of the Consistency and Wisdom indices with these variables are reported in Table 2. Only 2 of the 49 correlations are statistically significant and in the hypothesized direction, and one of these relationships is due to an artifact: Accessibility is correlated .42 with Socioeconomic Level, but the latter constitutes part of the former. The one presumably true relationship is that of Agreement between Measured Interests and Preferences with Patterning of Interests (.29): boys whose preferences agree with their measured interests tend to have clear-cut patterns of measured interests, a relationship which makes excellent psychological sense and which gives one a little confidence in measured interests as having some meaning at the ninth grade level, despite their lack of relationship to other variables. (But when interests are related only to interests one wonders if there is perhaps a measurement artifact at work here also.)

Five other correlations in Table 2 are large enough to be significant, but since they are in the unexpected direction and do not make psychological sense they must be attributed to chance. In a table of 49 correlations, one might find 2

Table 2: *Relationships of consistency and wisdom of ninth-grade vocational preferences to status and background variables*

	Age in Grade 9	Socioecon.. Level	Intell. in Gr. 9	School Achiev.	Pattern Interests	TAT Adj.	Peer Accept.
Consistency:							
Within Fields	—11	—04	—04	09	—.14	—04	—01
Within Levels	—08	—03	06	00	—.08	05	05
Within Families	—12	—02	02	00	—.18	03	—01
Wisdom in terms of:							
Aptitudes	—16	—07	12	03	—.27	—10	—12
Measured Interests	—12	02	00	06	.29**	14	07
Measured Interest Level	14	—11	—22	—06	—.27	—06	—17
Accessibility	12	42**	—01	02	—.09	—07	—02

Note.—N = 105 boys. Decimal points have been omitted.
** Significant at the .05 level, one-tailed test.*
*** Significant at the .01 level, one-tailed test.*

or 3 correlations which seem statistically significant at the .05 level, and 1 correlation significant at the .01 level, strictly on a chance basis.

The conclusion to be drawn from Table 2, as from Table 1, thus appears to be that the consistency and wisdom of vocational preferences have little significance for prevocational and vocational choices at the ninth grade level.

Implications

If the consistency and wisdom or realism of a ninth grade boy's expressed vocational preferences lack validity, as the Career Pattern Study data suggest, does this mean that they should be disregarded in practice? The fact that consistency and wisdom of vocational preferences have little meaning at Grade 9 does suggest that they should not be used as criteria of the need for guidance nor of the effectiveness of guidance at that stage of development. Presumably at this stage the very instability and transiency of expressed vocational preferences, and perhaps also their inadequate factual basis, make their apparent consistency and wisdom largely a matter of chance. However, one would be reluctant to draw the conclusion that the preferences themselves, even if inconsistent or unwise, should be disregarded, for this would involve not only the assumption that the counselor knows better than the pupil what is appropriate for him, but also the assumption that the best way to help the pupil to choose and plan wisely is to get him to concentrate on data coming from without himself rather than to examine his self-concept in relation to external and impersonal data. The demonstrated, even though far from perfect, validity of aptitude, achievement, and interest test data make it clear that the counselor does indeed have unique externally derived information as to what is appropriate for the pupil. But it may well be, as many counselors believe and as some research shows, that the best way to let this information help the pupil is to aid him in assimilating it into his concept of himself. This is best done by beginning, not with the data, but with the self-concept. And the statement of a vocational preference is one way of expressing a self-concept, as Bordin (1943) and the present writer (1951) have pointed out.

There are, of course, various ways of helping in the exploration of the self-concept as manifested in an expressed vocational preference. One is to help the pupil to choose courses, activities, part-time employment, etc. in which he may find opportunities to see if the preferred role does indeed suit him, and to discuss his use and evaluation of these experiences with him as he participates in them. Another is to help him to examine his preference in the interview, relating it to his picture of himself as a student, as a part-time worker, as a member of a family group, etc. In this discussion he may be helped to consider the views of him held by other persons and the picture of himself obtainable from records of his performance. This discussion can lead to the desirability of obtaining other such external data, and more testing or more participating in courses or activities may be planned and the results later reviewed by pupil and counselor. By thus using expressed vocational preferences as springboards for exploration and growth, for self-evaluation and further planning, the counselor proceeds democratically and with respect for his client's right to and need for self-determination, while avoiding the error of placing undue emphasis on preferences as a basis for directional choice.

SUMMARY

Questions are raised concerning the significance of the consistency and wisdom or realism of vocational preferences among ninth grade boys, and the use of measures of these as criteria of the effectiveness of or the need for vocational guidance. The development and application of several such measures in the Career Pattern Study are described, and data on the construct validity of these indices for 105 typical boys are reported. The failure to find significant relationships in the hypothesized directions is taken as evidence of the lack of psychological and hence of practical educational significance of consistency and wisdom or realism of vocational preferences at this stage of development. It is concluded that, although they may be meaningful at later stages, they should not be used as criteria of the need for or effectiveness of guidance and counseling at the ninth grade level.

REFERENCES

BORDIN, E. S. A theory of vocational interests as dynamic phenomena. *Educ. psychol. Measmt.*, 1943, *3*, 49–65.

CARTER, H. D. Vocational interests and job orientation. *Appl. psychol. Monogr.*, 1944, No. 2.

DARLEY, J. G. *Clinical application and interpretation of the Strong Vocational Interest Blank.* New York: Psychological Corp., 1941.

FRYER, D. *The measurement of interests.* New York: Holt, 1931.

HAMBURGER, M. Realism and consistency in early adolescent aspirations and expectations. Unpublished doctoral dissertation, Teachers College, Columbia University, 1958.

MOSER, HELEN P., DUBIN W., & SHELSKY, I. A proposed modification of the Roe Occupational Classification. *J. counsel. Psychol.*, 1956, *3*, 27–31.

ROTHNEY, J. W. M. *Guidance practices and results.* New York: Harper, 1958.

SCHMIDT, J. L., & ROTHNEY, J. W. M. Variability of vocational choices of high school students. *Personnel guid. J.*, 1955, *34*, 142–146.

STRONG, E. K., JR. *Vocational interests of men and women.* Stanford: Stanford University Press, 1943.

SUPER, D. E. Vocational adjustment: Implementing a self-concept. *Occupations*, 1951, *30*, 88–92.

SUPER, D. E. *The psychology of careers.* New York: Harper, 1957.

SUPER, D. E., CRITES, J. O., HUMMEL, R. C., MOSER, HELEN P., OVERSTREET, PHOEBE L., & WARNATH, C. F. *Vocational development: A framework for research.* New York: Bureau of Publications, Teachers College, Columbia University, 1957.

SUPER, D. E., & OVERSTREET, PHOEBE L. *The vocational maturity of ninth-grade boys.* New York: Bureau of Publications, Teachers College, Columbia University, 1960.

WARNER, W. L., MEEKER, MARCHIA, & EELLS, K. *Social class in America.* Chicago: Science Research Associates, 1949.

INTELLIGENCE AND CREATIVITY

56

Psychological Health and Cognitive Functioning in Adolescence: A Multivariate Analysis

Philip W. Jackson, Jacob W. Getzels, and George A. Xydis

The authors of this paper explore the relationships between age, sex, and the measures of psychological health and cognitive functioning. Their study demonstrates, as do other articles in this book, that such relationships are extremely complex. Getzels and Jackson have also studied adolescent intellectual functioning, using Guilford's (3) model of intelligence, in their book, Creativity and Intelligence *(1962).*

Naive conceptions regarding the relationship between psychological health and cognitive functioning pervade both the professional literature and the popular press. Most crudely stated, these conceptions imply that as psychological health goes, so goes cognition; that disorders and malfunctioning of the former will be accompanied invariably by disorders and malfunctioning of the latter. On the basis of such beliefs educators and other professional workers have been led to assume that any practice designed to promote a person's psychological well-being will produce direct and substantial gains in the cognitive acuity of that person.

When one seeks evidence to support this general conception, the results are disappointing. Both clinical and empirical investigations concerned with the interplay between psychological health and cognition attest to the complexity of this relationship. Rapaport (1945), for example, describes the dynamics by which intellectualization, or the libidinization of thought-processes, comes to be used as a potent defense mechanism for the individual. More recently, Haggard and others (1957) have identified particular personality constellations in children, some involving pathology, that are associated with outstanding achievement in certain scholastic areas. In an earlier paper the present authors described the differences in scholastic achievement which accompany public-private conflicts within the individual (1958). From these and other investigations it is apparent that differences in the definitions both of psychological health and of cognitive functioning produce corresponding differences in the conclusions which may be drawn concerning the relationship between the two. The need to specify the conditions which affect this relationship is evident.

PROBLEM

The research described here analyzes the effect of four major classes of variables upon the relationship between psychological health criteria and cognitive performance. These variables include: (a) five measures of psychological health, (b) eight types of cognitive performance, (c) six school grade levels, and (d) sex. It is the purpose of this research to study the contribution each of these variables makes to observed variability in the psychological health-cognition relationship. In addition, the interaction among these classes of variables is examined.

SUBJECTS, INSTRUMENTS, AND PROCEDURES

The subjects of this research were 292 boys and 241 girls enrolled in a Midwestern private school. These students were divided into six class groups ranging from the sixth grade to the senior year of high school. In this institution a single grade, the "pre-freshman," is substituted for the usual seventh and eighth grade levels so that the high school group is one year advanced when compared with most public school adolescents. All subjects were given the following tests of cognitive ability:

1. *Individual intelligence tests.* In most cases, this was the WISC. A small number of children were given the Henmon-Nelson, the

---

From *Child Development*, 1960, *31*: 285–293. Reprinted with permission of the Society for Research in Child Development.

scores of which were converted by regression equation into equivalent WISC scores.

2. *Standardized numerical achievement test.* Due to differences in the curricula of various grade groups it was not possible to administer the same test of numerical achievement to all subjects. The following tests were given according to grade placement:

Sixth Grade: Iowa Everypupil Arithmetic Test, Advanced Form O
Pre-Freshmen: Snader General Mathematics Test
Freshmen: Cooperative Elementary Algebra Test, Form T
Sophomores: Cooperative Intermediate Algebra Test
Juniors: Cooperative Geometry Test, Form 2
Seniors: Cooperative Geometry Test, Form 2

3. *Standardized verbal achievement test.* The Cooperative Reading Test was given to all students from the sixth grade through the junior year of high school. Sixth graders and pre-freshmen were given Test C1, Form Y; the remaining three groups received Test C2, Form T of the Cooperative Reading Instrument. The verbal achievement of seniors was assessed by means of the USAFI GED Test 1.

4. Five "creativity" measures adapted from the test batteries of Guilford (1954) and Cattell (1955), or designed especially for this study, as follows:

a. *Word Association.* The subject was asked to give as many definitions as possible to fairly common stimulus-words, e.g., "bolt," "bark," "sack." His score depended upon the absolute number of definitions and the number of different categories into which these definitions could be put.
b. *Uses for Things.* The subject was required to give as many uses as he could for objects that customarily have a stereotyped function attached to them, e.g., "brick," "paper clip." His score depended upon both the number and originality of the uses which he mentioned.
c. *Hidden Shapes.* The subject was required to find a given geometric form that was hidden in more complex geometric forms or patterns.
d. *Fables.* The subject was presented with four fables in which the last lines were missing. He was required to compose three different endings for *each* fable: a moralistic, a humorous, and a sad ending. His score depended upon the number, appropriateness, and originality of the endings.
e. *Make Up Problems.* The subject was presented with four complex paragraphs each of which contained a number of numerical statements, e.g., "the costs involved in building a house." He was required to make up as many mathematical problems as he could that might be solved with the information given. His score depended upon the number, appropriateness, and originality of the problems.

The following psychological health instruments were administered to all subjects:

1. *Group Rorschach.* Cards III, IV, IX, and X were projected on a screen. For each picture the subject was presented with 10 responses and was asked to choose the three which he thought to be most appropriate. Each list of 10 contained four "pathological" responses. The subject's score was the number of nonpathological responses among his 12 choices.

2. *Direct Sentence Completion.* The subject was asked to complete 27 sentences of the type: "When I saw I was going to fail I ———," or, "I think my father is ———." Each sentence was given a plus or minus score depending upon the presence or absence of morbid fantasy, defeatism, overt aggression, and the like. The total score was the summation of the individual sentence scores.

3. *Indirect Sentence Completion.* This instrument was identical with the Direct Sentence Completion except that proper names were inserted for the pronoun "I," thus changing it from a "self-report" to a "projective" instrument. Boys' names were used in the male form of the instrument and girls' names in the female form. The instrument was presented as a "thinking speed" test. To reinforce this notion subjects were asked to raise their hands when they were finished and the elapsed time was written on the front of their test booklet. This instrument was administered approximately two weeks *prior* to the administration of the Direct Sentence Completion.

4. *California Personality Test.* Two forms of this instrument were used. The intermediate form was given to sixth graders and pre-freshmen; the secondary form was given to all of the older groups. Two subscores were obtained, one in "personal adjustment" and one in "social adjustment."

Correlation coefficients were computed for all cognitive-psychological health combinations. The data were treated separately by sex and by school grade so that at each of the six grade levels 80 coefficients were produced, 40 for each sex. In order to test for the effects of sex, school-grade, and sex-grade interactions upon the variability of the correlations, the coefficients were transformed into Fisher z scores and standard chi square statistics were computed. It was not possible to follow the same procedure in the analysis of main effects for the psychological health and cognitive variables since the assumption of independence of observations (correlation coefficients) could not be met. In this latter section of the analysis tests of statistical significance were limited to selected combinations of variables.

RESULTS

Considering the total population of correlation coefficients, the relationship between psychological health and cognition appears to be rather weak, although quite variable. The mean of the correlations was .04, and the range was from −.56 to +.56. As the following analysis will show, however, it is incorrect to subscribe to the generalization that little or no relationship exists between psychological health criteria and cognitive functioning. When the effects of the four major variables are examined, the relationship is seen to alter in a systematic fashion.

Chi square statistics computed to test sex, grade, and sex-grade interaction effects among the correlation coefficients are given in Table 1. As this table indicates, the relationship between

Table 1: *Chi square statistics computed to test sex, grade, and sex-grade interaction effects among correlation coefficients*

		Psychological Health Variable				
Cognitive Variable	*Sex-Grade Variable*	*Group Rorschach*	*Direct Sentence Completion*	*Indirect Sentence Completion*	*California Personal Adjustment*	*California Social Adjustment*
Fables	Sex	—	—	—	—	—
	Grade	—	—	—	—	—
	Sex-Grade	—	—	—	—	—
Uses	Sex	4.52*	—	—	—	—
	Grade	12.96*	—	—	—	—
	Sex-Grade	—	16.75***	—	—	—
Hidden Shapes	Sex	3.98*	—	—	—	3.84*
	Grade	11.30*	—	—	—	—
	Sex-Grade	—	—	—	—	—
Word Association	Sex	5.49**	—	—	—	—
	Grade	19.27***	—	—	—	—
	Sex-Grade	—	—	—	—	—
Make Up Problems	Sex	—	5.19*	—	—	7.82***
	Grade	—	—	—	—	—
	Sex-Grade	—	—	—	—	—
Verbal Achievement	Sex	9.31***	—	—	—	—
	Grade	—	—	—	—	11.71*
	Sex-Grade	—	11.77*	13.94**	—	—
Numerical Achievement	Sex	—	—	—	—	—
	Grade	—	11.75*	—	—	11.93*
	Sex-Grade	—	—	—	—	—
IQ	Sex	5.34*	—	—	—	—
	Grade	—	—	—	—	—
	Sex-Grade	14.19**	—	—	—	—

Note.—All correlations were transformed to Fisher z scores before chi squares were computed. Chi squares with a probability level less than .95 are not shown in this table.

* *Significant at the .05 level.*

** *Significant at the .02 level.*

*** *Significant at the .01 level.*

Fig. 1: *Plots of correlation coefficients showing significant changes at various grade levels—boys.*

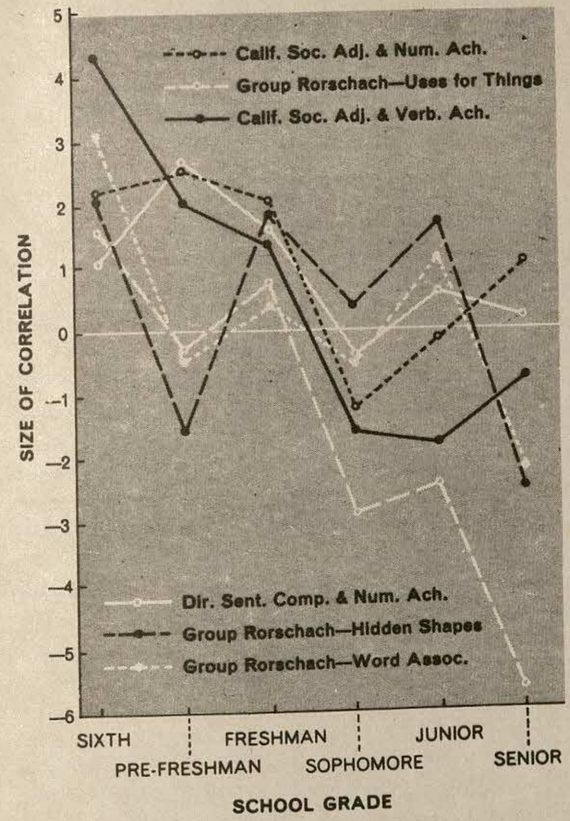
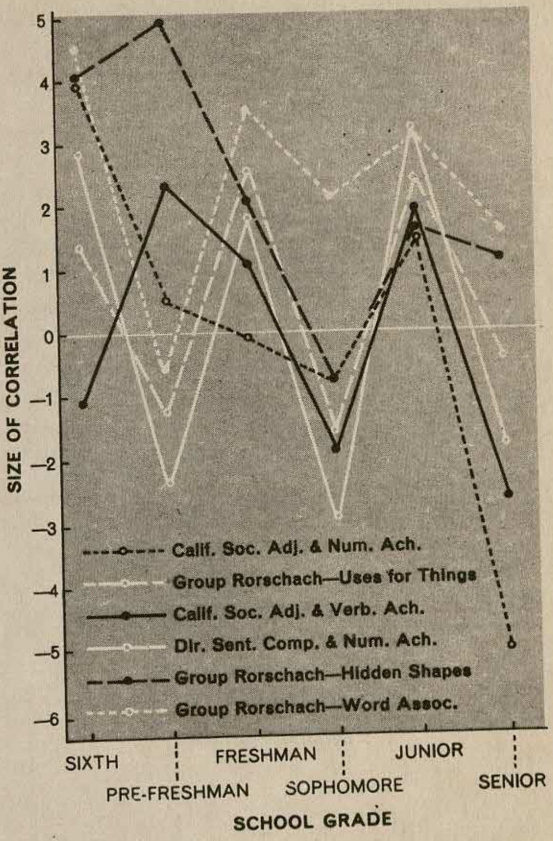


Fig. 2: *Plots of correlation coefficients showing significant changes at various grade levels—girls.*



psychological health and cognition, as measured by specific criterion variables, reflects both sex and grade differences. Only two of the instruments used in the study fail to show significant variability with sex, grade, or sex-grade interaction. These two variables are the California Personal Adjustment score and the Fables Test. It is also important to note that the effects of sex and grade are far from uniform throughout the table. Thus, depending on the particular pair of variables chosen, one could find support for almost any hypothesis dealing with the effect of sex and grade variables.

The single psychological health variable that seems most sensitive to sex and grade variability is the Group Rorschach. Nine of the 18 significant chi squares in Table 1 involve this variable. A more precise description of this variability will be given below.

The correlation coefficients which show significant variability with grade are presented graphically in Figures 1 and 2. The sexes are presented separately since some of the grade differences are also combined with differences between boys and girls.

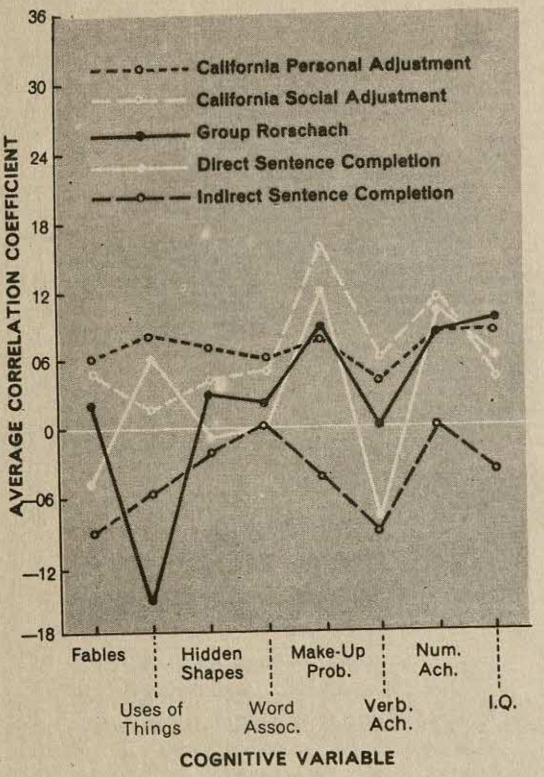
The following observations concerning these two graphs seem appropriate:

1. For boys there is a general decline in the relationship with age. The sixth grade period is the one during which the relationship between psychological health and the quality of cognitive performance appears to be highest. In addition, there is a tendency for the variability among these correlations to become greater as one moves from the lower to the higher grades.

2. Among senior boys there is a pronounced negative relationship between the Group Rorschach and three tests of creativity.

3. The correlation coefficients for girls are much more cyclic in their variability. They are highest, on the average, and much less variable in the junior year of high school.

Fig. 3: *Average correlations between psychological health variables and cognitive variables—boys.*



4. Unlike senior boys, the senior girls show pronounced negative correlations between performance on academic tests and psychological health measures involving self-report or personal appraisal.

When school grade differences are removed by a procedure of averaging, the correlation coefficients yield the general patterns shown in Figures 3 and 4.

The major findings shown in Figures 3 and 4 may be summarized as follows:

1. The meaning of the sex differences associated with the Group Rorschach (shown in Table 1) becomes dramatically apparent in these two graphs. This instrument shows a striking relationship to cognitive functioning among girls, but not among boys. In fact, no other psychological health variable in either figure approaches the degree of relationship between pathological fantasy, as measured by the Group Rorschach, and cognitive functioning. Further, the coefficients relating the Group Rorschach to cognitive variables increase noticeably as one moves from creativity tests to measures of scholastic achievement and IQ.

2. For both groups the type of cognitive functioning which on the average shows the greatest sensitivity to psychological health is that which involves the manipulation of numerical symbols. This tendency is somewhat more pronounced in the boys' data.

3. For girls the great majority of the relationships are negative (with the obvious exception of the Group Rorschach); while for boys the majority of the coefficients are positive.

DISCUSSION

The extreme range and average size of the correlation coefficients treated in this investigation point to the dangers inherent in making broad and sweeping statements concerning the relationship between psychological health and cognitive functioning in adolescence. These data make clear the necessity of qualifying such statements. More specifically, they demonstrate that sex, school grade, the type of psychological health criterion, and the type of cognitive functioning are among the variables that must be considered when making these qualifications.

Among all of the findings described above, those which concern sex differences and the role of pathological fantasy seem to us most worthy of special comment.

Succinctly stated, the above data suggest that the cognitive functioning of girls is more sensitive to aberrations of fantasy than is the cognitive functioning of boys. Also, it should be mentioned that no differences exist among boys and girls in the relative frequency of pathologic responses. Thus, although girls are as likely to choose pathological responses as are boys, this choice gives more information about the cognitive performance of the former than of the latter group.

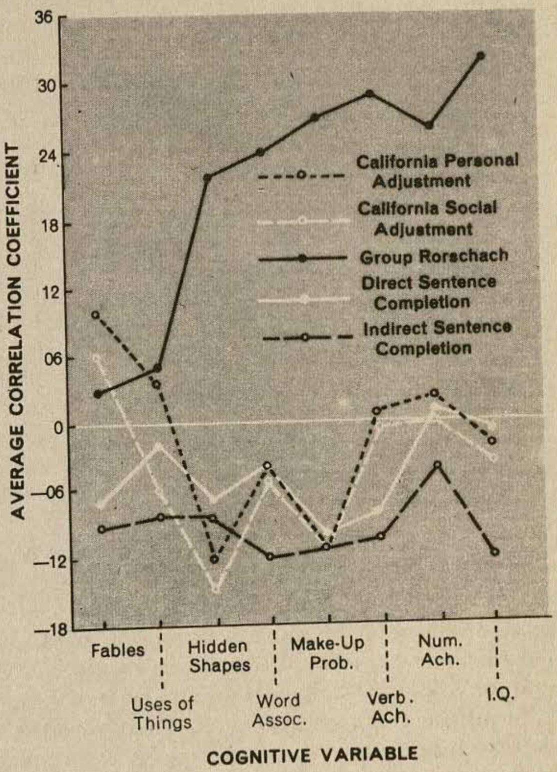
A tentative explanation of this sex difference might focus on the nature of the responses to the Group Rorschach and the relationship of these responses to differences in cultural expectations for boys and girls.

Among the 16 choices which were considered to reflect pathology on the Rorschach, a great number appear to involve *morbidly aggressive* fantasy, e.g., "bloody stomach," "a burnt mass," "blood and dirt," "an explosion," "a dirty nasty mess," "bloody clouds," etc. Both common observation and recent studies support the notion that society tends to condone, if not encourage, the aggressive actions of boys, while discouraging, if not condemning, similar action on the part of girls. Given these social sanctions aimed at overt behavior, it is not far-fetched to hypothesize correlative self-imposed prohibitions referring to the inner world of the person. Although both sexes may think an equal number of "bad thoughts," the threat to the individual's concept of self-worth instigated by these thoughts may be far from equal for boys and girls. If morbid and aggressive fantasies tend to be less threatening to boys, their occurrence would be less likely to set off a chain of introspective events interfering with rational cognitive functioning. Girls, on the other hand, might be less capable of ignoring this "internal disorder," and accordingly there would be greater interference with rational cognitive functioning.

The relationship shown in this research between pathological fantasy and different types of cognitive performance is quite consistent with both traditional and recent formulations concerning the psychological functioning of creative people. The suggestion that fantasy plays an important part in the creative act has been a matter of record for centuries. At times irrational fantasy itself has been the subject of artistic effort as in *Kubla Khan* or the surrealist's landscape. More recently the relationship between fantasy and creativity has become the focus of psychological investigation, and empirical evidence supporting the intuitive insight of artists and poets is steadily increasing. If it can be assumed that the presence of atypical fantasy is also a rough indicator of a general *richness* of fantasy life, then the relative decline in the relationship between the Rorschach and cognitive performance as one moves from scholastic performance to tests of creativity becomes more understandable. Since the tests of creativity call for original or unconventional responses, rather than "right" answers, it would follow that the individual who possesses a rich (and perhaps, at times, bizarre) fantasy life would be at an advantage. Conversely, the absence of an extremely rich fantasy life might be somewhat of a blessing when confronted with cognitive tasks which require intensive and prolonged absorption in a highly rational system.

The meaning of school-grade differences in the relationship between psychological health and cognition remains rather obscure. Grossly interpreted, the findings suggest that psychological difficulties intrude into the cognitive performance of boys during the pre- or early adolescent period while the same phenomenon among girls does not occur until the junior year of high school. Since the general level of scores on the psychological health instruments shows no consistent age pattern, one can only assume that it is the quality rather than the quantity of the problems facing the students during these two periods that is important. Our general knowledge concerning the developmental problems facing adolescents during various periods suggests that the early adolescent period for boys is one in which physical maturity becomes crucial. It is during this period that general variability in size, strength, and the possession of secondary sex characteristics becomes most prominent. A tentative hypothesis would be that concern over delayed or extremely premature development during this period accounts for the general rise in the relationship between psychological health and cognition.

Fig. 4: *Average correlations between psychological health variables and cognitive variables—girls.*



Girls, on the other hand, face an equally vexing set of problems in middle adolescence. These problems, rather than focusing on somatic changes, have as their general theme the resolution of social relationships. It is during this period that "chum" relations and involvement with "the crowd" gradually lose their appeal in favor of the more serious business of dating and "going steady." The demands which these more intimate heterosexual relations make upon the psychological well-being of girls may serve to explain the increased relationship between psychological health and cognition during this period.

SUMMARY

The relationship between psychological health and cognition in adolescence was examined using five psychological health criteria and eight types of cognitive performance. The subjects were 292 boys and 241 girls enrolled in a Midwestern private school.

The major findings indicated that the relationship under question shows systematic variability with school grade, sex, psychological health criteria, and type of cognitive function. In particular, the interrelationship between sex differences and the role of pathological fantasy pose a series of promising research questions. A tentative theoretical framework was suggested for a number of the major findings.

REFERENCES

CATTELL, R. B. *Handbook for the objective-analytic personality test batteries*. Champaign, Ill.: Inst. Personality & Ability Testing, 1955.

GUILFORD, J. P., KETTNER, N. W., & CHRISTENSEN, P. R. A factor-analytic study across the domains of reasoning, creativity, and evaluation: I. Hypotheses and description of tests. *Rep. Psychol. Lab., University of Southern California*, 1954, No. 11.

HAGGARD, E. A. Socialization, personality, and academic achievement in gifted children. *Sch. Rev.*, 1957, *65*, 388–414.

JACKSON, P. W., & GETZELS, J. W. Psychological health criteria among adolescents: A study of manifest and latent indicators. Paper read at Amer. Psychol. Ass., Washington, D.C., September 1958.

RAPAPORT, D. *Diagnostic psychological testing*. Vol. I. Chicago: Yearbook Publishers, 1945.

57

The Identification of Creativity in Adolescents

Ellen V. Piers, Jacqueline M. Daniels, and John F. Quackenbush

The Guilford model (3) has also been used by Piers and her colleagues to explore creativity in adolescents. Many of the current studies in this field use divergent thinking measures adopted from Guilford. The problem of an acceptable definition for creativity still plagues this rather new area of research, but empirical studies, such as this one, assist in clarification of both the concept and its measurement.

The subject of creativity, so long neglected by psychologists, is currently being revived. Some investigators, Barron (1958), for example, are working with prominent adults in a variety of fields to try to determine aspects of the creative personality. Others (Guilford, Kettner, & Christensen, 1954, 1956) are using factor analytic methods in an attempt to validate the construct and explore its generality. Guilford and his co-workers have devised some novel types of tests (most of which are still unpublished) from which promising factors have been identified. Two factors which have emerged consistently, and which Guilford considers to be most useful in identifying persons with creative talent, are Originality, currently defined as the production of uncommon, clever, or remote responses, and Ideational Fluency, defined as the speed of calling up ideas, independent of their quality, in a situation in which there is relatively little restriction. Guilford places more weight on Originality.

Drevdahl (1956), in a study with college students rated by faculty members on a seven-point scale of creativity, found that Guilford's Originality was one of the measures which significantly differentiated those above and below the mean on the creativity rating.

Guilford (1950) has stated that creativity is more than intelligence, and cannot be accounted for adequately in terms of IQ. He has also stated that one of the most important aspects of the problem is the discovery of creative promise in our children and youth.

From *Journal of Educational Psychology*, 1960, *51*: 346–351

The present study is an attempt (*a*) to assess the usefulness of some of Guilford's experimental tests with a younger age group, (*b*) to report sex and grade differences on these tests, and (*c*) to determine the degree to which the tests agree with teacher ratings of creativity.

Getzels and Jackson (1958) attempted to differentiate the highly creative from the highly intelligent adolescent by means of five creativity measures taken or adapted from Guilford and Cattell, and then to compare the two groups on various measures. In the present study, however, all of the sample will be of above average intelligence, in order to try to distinguish differences in creativity within this group.

Table 1: *Age and IQ of junior high sample according to sex and grade*

	N	Mean Age in Mo.	*SD*	Mean IQ	*SD*
Grade 7					
Males	31	148.2	3.2	119.4	
Females	24	147.8	4.1	118.2	
Total	55	148.0	3.6	118.9	7.5
Grade 8					
Males	21	160.1	4.2	132.7	
Females	33	159.4	4.0	136.3	
Total	54	159.7	4.1	134.9	10.0
Total Sample					
Males	52	153.0	6.9	124.8	
Females	57	154.5	7.0	128.8	
Total	109[a]	153.8	7.0	126.9	11.9

[a] 5 cases not included in these calculations because of incomplete information.

METHOD

Subjects

*S*s were 114 seventh and eighth grade students from a junior high school. They ranged in age from 11 years, 2 months to 14 years, 1 month. Fifty-seven of the *S*s constituted the two top seventh grade classes and 57 constituted the two top eighth grade classes, which had been ability grouped by a combined ranking on intelligence and achievement test scores. Recent IQs from the Otis Intelligence Test were available from the school records of 110 *S*s. Table 1 shows the mean ages and IQs of the sample according to sex and grade. The discrepancy between the mean IQ of the seventh and eighth grades could not be explained by the school.

Measures

Guilford Tests of Creativity. Brick Uses requires the *S* to give as many uses for a brick as he can in 10 minutes. An ideational fluency score is obtained from the number of acceptable responses. Consequences requires the *S* to give as many answers as he can in 2 minutes to each of four questions, such as "What would happen if all books were destroyed?" Direct or immediate consequences constitute the low quality or fluency score, while indirect consequences constitute the remoteness or originality score. Plot Titles requires that the *S* give as many titles as he can in 3 minutes to each of four short plots. Each response is rated as "clever" or "nonclever" by judges. Nonclever responses constitute the low quality or ideational fluency score, while clever responses constitute the originality score. Quick Responses requires the *S* to respond with the first word that comes to his mind, to a list of 50 words read at the rate of 5 seconds per word. An originality score is obtained by ranking the response words in terms of their frequencies and assigning weights to each word, then summing the weights for all of the *S*'s responses. Unusual Uses requires the *S* to give as many unusual uses as he can for each of six common objects, in two 5-minute periods. An originality score is derived by summing the number of acceptable responses.

Teacher Rating Scale. A five-point rating scale ranging from "extremely creative" to "extremely uncreative" was developed to obtain teacher judgments. In order to establish the same frame of reference among the teachers, definitions of each of the five rating categories and a list of several examples of behavior, which (according to the theories of Guilford and others) would fit the criterion, were given to them with the rating scale, along with the following definition adapted for purposes of this study:

> Creativity is the capacity of the individual to avoid the usual routine, conventional ways of thinking and of doing things and to produce a quantity of ideas and/or products which are original, novel, or uncommon and which are workable. It must be purposeful or goal directed. It may involve the forming of new patterns and combinations of information derived from past

Table 2: *Means and standard deviations of scores on Guilford Tests for two samples*

	Guilford Sample				Junior High Sample (N = 114)		
	N	Mean	*SD*	*Reliability*[a]	Mean	*SD*	*Reliability*[b]
Consequences							
Remoteness score	410	10.3	4.7	.66	7.3	4.2	.68
Low Quality score	410	14.9	6.2	.68	13.5	4.3	.50
Plot Titles							
Cleverness score	364	6.4	3.7	.43	3.4	2.8	.73
Low Quality score	364	12.2	6.9	.70	19.0	7.9	.90
Brick Fluency	—	—	—	—	19.0	6.3	—
Quick Responses	410	99.8	18.7	.81	85.8	23.4	.84
Unusual Uses	204	15.4	5.2	.68	10.8	3.8	.66

[a] *All alternate form estimates except Quick Responses, which is an odd-even estimate.*
[b] *All odd-even estimates.*

Table 3: *Grade differences on Guilford Tests and teacher ratings* (N=114)

	Grade 7			Grade 8		
	Mean	*SD*	Reliability	Mean	*SD*	Reliability
Consequences						
Remoteness score	6.0	3.8	.62	8.4	3.8	.68
Low Quality score	13.4	4.4	.59	13.6	4.5	.40
Plot Titles						
Cleverness score	2.8	2.2	.56	4.0	2.1	.78
Low Quality score	17.7	5.8	.87	20.2	7.0	.92
Brick Fluency	18.1	5.3	—	19.8	7.0	—
Quick Responses	82.2	23.1	.82	89.3	23.7	.86
Unusual Uses	9.4	3.2	.45	12.0	4.0	.73
Teacher Ratings	6.1	1.6		6.2	1.9	

experience, and the transplanting of old relationships to new situations, or the generation of new relationships.

Procedure

Each of the four classes of *S*s was tested for two 50-minute periods on the same day, in order to prevent out-of-class discussion. Each class was tested one period by each of the two examiners. Teachers were not present. With a few exceptions, the *S*s took the study seriously, and appeared to do their best.

Art, mathematics, English, social science, and science were considered the most appropriate classes from which to select the teachers. Most of the teachers chosen taught at least two of the areas, so that three teacher ratings were obtained for each *S*. These ratings were summed and coded for ease in computation. The coded score was the Teacher Rating.

RESULTS

In an effort to increase the reliability of the scoring, each of Guilford's tests was scored jointly by two of the investigators, and in the case of Quick Responses and Plot Titles, by all three investigators.

Seven scores were obtained from these tests. Unusual Uses, Quick Responses, Remote Consequences, and Clever Plot Titles represented the Originality factor. Brick Uses (Fluency), Immediate Consequences, and Nonclever Plot Titles represented the Ideational Fluency factor.

Table 2 shows the means and standard deviations for the creativity tests for this sample, as

Table 4: *Correlations between teacher ratings of creativity*

Grade	*N*			
7-1	24	$\tau_{AB} = .10$	$\tau_{AC} = -.14$	$\tau_{BC} = .23$
7-8	31	$\tau_{BD} = .31$	$\tau_{BE} = .62$*	$\tau_{DE} = .23$
8-3	28	$\tau_{FG} = .16$	$\tau_{FA} = .20$	$\tau_{GA} = .13$
8-8	29	$\tau_{HI} = .34$	$\tau_{HG} = .28$	$\tau_{IG} = .34$

* *Significant at the .01 level.*

Table 5: *Sex differences on Guilford Tests and teacher ratings*

	GRADE 7				GRADE 8				TOTAL			
	Male		Female		Male		Female		Male		Female	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Consequences												
Remoteness score	5.6	3.7	6.7	3.8	9.0	4.3	8.2	4.4	7.0	4.3	7.6	4.2
Low Quality score	13.4	4.9	13.3	3.8	13.6	4.4	13.3	3.8	13.5	4.7	13.3	3.8
Plot Titles												
Cleverness score	2.7	2.0	2.8	2.3	3.3	3.2	4.6	3.3	2.9	2.6	3.8	3.0
Low Quality score	15.6	6.5	20.5	6.5	21.6	8.2	20.0	8.7	18.0	7.8	20.0	7.8
Brick Fluency	16.7	4.8	19.8	5.5	18.8	5.9	20.1	7.6	17.6	5.3	20.0*	6.8
Quick Responses	80.4	17.6	84.6	28.6	92.7	20.4	87.4	25.0	85.4	19.7	86.2	26.5
Unusual Uses	9.7	3.4	9.2	3.0	11.4	4.1	12.4	3.9	10.4	3.8	11.0	3.9
Teacher Ratings	6.0	1.7	6.2	1.5	6.6	1.6	6.1	1.9	6.3	1.7	6.1	1.8

* *Significantly different at .05 level.*

compared with one sample of Air Cadets used by Guilford. Inspection of the table reveals substantially the same types of dispersion for the two samples. As would be expected, the younger age groups are not quite as productive, except for the low quality (nonclever) Plot Titles.

Since alternate forms of the tests were not available, and since most of the tests were made up of separately timed units of approximately equal difficulty, odd-even estimates of reliability were computed, using the Spearman-Brown formula. Table 2 shows that the reliabilities for this sample compare favorably with Guilford's sample, except in the case of Immediate (low quality) Consequences.

Table 3 shows the Junior High sample broken down into the two grade levels. As would be expected, the eighth grade group is slightly higher for all measures, but results are sufficiently similar to justify combining the two grades.

Correlations were obtained for each pair of the three teachers who rated each grade. In spite of the efforts made to define the criterion, Table 4 shows that teachers do not agree very well in rating students on degree of creativity. Only one correlation proved to be significantly different from zero. While it is evident that teachers are using different criteria for their ratings, their judgments were pooled in order to broaden the bases upon which scores for creativity were estimated.

In appraising creativity, sex differences may also be important. Table 5 shows, however, that in this sample, only Brick Fluency showed a consistent difference in favor of the girls, the difference being significant beyond the .05 level for the total group.

Raw scores on the creativity tests were correlated with each other, with Teacher Ratings and with IQs. Since no ratings were available for two *Ss*, all correlations involving Teacher Rating were based on 112 *Ss*. Results are shown in Table 6, with the four tests purporting to measure Originality listed first. It can be seen that these four not only intercorrelate positively with each other, but they are also positively correlated with IQ, and, in two cases, with Teacher Ratings. The three Ideational Fluency tests, on the other hand, show no significant relationship either with Teacher Ratings or IQ.

Table 6: *Correlations of creativity tests with each other, with teacher ratings, and with IQs*

Tests	Quick Resp.	Conseq. Remote	Plot T Clever	Brick Fluen.	Conseq. Immed.	Plot T Noncl.	Teacher Ratings	IQ
Unusual Uses	.26**	.45**	.36**	.36**	.14	.28**	.07	.32**
Quick Responses		.28**	.49**	.03	.01	.05	.29**	.27**
Conseq. Rem.			.40**	.42**	—.12	.33**	.05	.31**
Plot Titles Clever				.25**	.27**	.18	.28**	.34**
Brick Fluency					.37**	.46**	—.21	.01
Conseq. Immed.						.24	—.002	—.04
Plot T. Nonclever							.14	—.06
Teacher Ratings								.20

*** Significant at the .01 level.*

Table 7: *Correlation of standard total scores on Guilford factors with teacher ratings* (N = *112*)

	Originality	Teacher Ratings
Fluency	.22*	—.02
Originality		.23*
Creativity		.16

** Significant at the .05 level.*

When the raw scores were changed to standard scores and combined into Originality and Ideational Fluency indices, as shown in Table 7, only Originality showed a significant correlation with Teacher Ratings, although the two indices were moderately correlated with each other. When all scores were combined into a Creativity Index, no significant correlation resulted.

DISCUSSION

One of the questions to be investigated was the degree to which Guilford's tests were suitable for a younger age group. Table 2 showed that the means and dispersions were very similar to the Air Cadet sample, and the reliabilities were, in general, of the same order. It would seem, therefore, that the tests, even in their present experimental form, are appropriate for research with bright junior high school students.

The second question had to do with sex and grade differences. In general, it appears that these tests do not discriminate markedly between the sexes. On only one test (Brick Fluency) was the difference consistently and significantly in favor of the girls. Mean grade scores showed small absolute differences in favor of the eighth grade on all tests. Since age variability within each grade was very small, and the mean difference between grades 11.7 months, the differences might be attributed to age, especially since Guilford's older Air Cadets showed higher scores on all tests reported except Low Quality Plot Titles. Certainly there appears to be an increase with age on all the Originality tests. However, the discrepancy in IQs between seventh and eighth grades must also be considered as a possible contributing factor, especially since the Originality tests were all positively related to IQ.

The third question concerned the empirical validation of the tests by means of Teacher Ratings. Results were disappointing and illustrative of the difficulties involved in this area. Partly due to the restriction of range within the sample, but also undoubtedly due to the vagueness and variability in the popular conception of what is meant by creativity, correlations between the three teacher ratings for each *S* were low. While pooled ratings represent heterogeneous points of view regarding creativity, they did correlate significantly with two of the Originality tests.

It would seem that we can look only for gross discriminations in teacher ratings of creativity at this age, if behavior in general is to be considered. This is not too surprising when we remember that teacher ratings of intelligence, originally used as a criterion for validating the Binet-Simon tests, are now considered to be less accurate than intelligence tests themselves. Validation of the Guilford tests may have to await longitudinal or follow-up studies in which scores are correlated with actual creative productivity in adult life, unless better methods can be devised of rating present performance or products.

As Guilford has hypothesized,[1] the tests which have factor loadings on Originality appear to be more satisfactory in the identification of creative talent than do those which have loadings on Ideational Fluency. The positive but only moderately high correlations of Originality with IQ also confirm his statement that creativity is more than intelligence. Getzels and Jackson (1958) have also found this to be true.

While the Guilford tests seem to be a promising method of identification of creative talent, a good deal more research is needed on these and other current instruments to establish both their concurrent and their predictive validity.

SUMMARY

Seven of Guilford's creativity tests, four with factor loadings on Originality and three with factor loadings on Ideational Fluency, were administered to 114 seventh and eighth grade students of above average intelligence and school achievement, and results were compared with teacher ratings of their creativity. Dispersions and reliabilities were found to be comparable to those of Guilford's Air Cadets, with some suggestion that there is an increase in mean score with age. Tests of Originality correlated more highly with IQ and with Teacher Ratings than did tests of Ideational Fluency. Teacher ratings of creativity proved to be a not very consistent criterion for validating the tests, due both to the restricted range and to vagueness of the concept of creativity. Sex was not a significant variable on most of the measures.

REFERENCES

BARRON, F. The psychology of imagination. *Scient. Amer.*, 1958, Sept., 150–166.

DREVDAHL, J. E. Factors of importance for creativity. *J. clin. Psychol.*, 1956, *12*, 21–26.

GETZELS, J. W., & JACKSON, P. W. The highly creative and the highly intelligent adolescent: An attempt at differentiation. Paper read at American Psychological Association, Washington, D.C., August 1958.

GUILFORD, J. P. Creativity. *Amer. Psychologist*, 1950, *9*, 444–454.

GUILFORD, J. P., KETTNER, N. W., & CHRISTENSEN, P. R. A factor-analytic study across the domains of reasoning, creativity, and evaluation: I. Hypothesis and description of tests. *Rep. Psychol. Lab., University of Southern California*, 1954, No. 11.

GUILFORD, J. P., KETTNER, N. W., & CHRISTENSEN, P. R. A factor-analytic study across the domains of reasoning, creativity, and evaluation: II. Administration of tests and analysis of results. *Rep. Psychol. Lab., University of Southern California*, 1956, No. 16.

---

[1] Personal communication to senior author.

58

Intelligence at Adulthood: A Twenty-Five Year Follow-Up

Katherine P. Bradway and Clare W. Thompson

Using the Stanford-Binet and the Wechsler Adult Intelligence Scale tests, Bradway and Thompson present a careful assessment of the intelligence of adults for whom data at two earlier ages was available. Again, the complex nature of both the problem of individual predictability through time and the concept of intelligence itself is illustrated. The authors challenge the notions of a fixed intelligence position in relation to one's age group and of a ceiling on intellectual development. See Bayley (4) for another type of longitudinal study.

Attempts to understand human intelligence were based initially on studies of children. In contrast, this paper deals with aspects of intelligence which cannot be studied adequately from data limited to childhood intelligence. Specifically, our approach has been to study as young adults a group of individuals previously tested during the preschool years and during adolescence.

Binet and Simon developed the original intelligence test to objectify the criteria which educators had been using for many years in assessing the capacities of children for purposes of school placement. The several early American adapta-

---

From *Journal of Educational Psychology*, 1962, *53*: 1–14.

tions of this test culminated in the 1916 Stanford-Binet, which provided indices of adult intelligence. The standardization group, however, consisted largely of children; adult indices were based on extrapolation.

The urgency of the needs for screening and assignment created by World War I led to the creation of truly adult intelligence tests but, consistent with these needs, these tests were for group rather than individual administration. Later revisions of individual intelligence tests (e.g., the 1937 and 1960 Stanford and the 1939 Kuhlmann) extended the upper limit by providing more difficult test items. Moreover, somewhat older subjects were included in the standardization groups, although the upper age levels of these groups did not go above the adolescent years into adulthood.

No adequate instrument existed by which the clinician could measure the intelligence of the individual adult until the publication of the Wechsler-Bellevue in 1939. Wechsler's large standardization sample started with children of 12 years and included older age groups up to 65 years. His emphasis on adult intelligence resulted in departure from the use of mental age, a concept which had contributed immeasurably to the interpretation of childhood intelligence, but which introduced confusion in the assessment of adult intelligence. His term "IQ" is not a continuation of Stern's quotient of mental age divided by chronological age; it is more comparable to a standard score. Unlike the revisions of the Binet, all of which had been based on the concept that intellectual abilities reach their peak at from 14 to 16 years of age, indices of intelligence for the first time took account of the continuation of intellectual growth into the adult years. This view of intellectual growth has been increasingly confirmed by more recent studies (Bayley, 1955; Bayley & Oden, 1955) including the present one. Although Wechsler calls his concept of intelligence global, the structure of the Wechsler-Bellevue (and its later revision, the WAIS) is such that several kinds of tasks are separately measured and separate indices of verbal and nonverbal abilities are provided.

The clinician trying to answer questions about the intelligence of an adult needs information about the unique growth rates of different kinds of abilities. No cross-sectional sampling can provide such information accurately. On the other hand, longitudinal studies pose problems of practice effects and of the attrition which inevitably occurs over a period of time. The clinician is also confronted with the problem of how most meaningfully to express the measured intelligence of any given adult. Should it be an index which compares him with his own age group? One which locates him on his own course of development and decline? Or several which represent an assessment of his various kinds of mental processes? Moreover, each of these requires that procedures for equating means and variability be worked out. Answers to many of these questions are necessary before valid predictions about adult intelligence can be made.

PROCEDURE AND SUBJECTS

Information for some of these answers is provided by the longitudinal study we have recently completed (Bradway & Robinson, 1961). A group of 111 adults aged 27 to 32 who had been tested 25 years earlier as part of the 1937 Stanford-Binet standardization sample of preschool children and again during their adolescent years was administered the Stanford-Binet and WAIS in 1956. The repetition of the Stanford-Binet permits analysis of improvement in the same tasks from adolescence to adulthood; the WAIS data permit comparison of child and adult measures of intelligence of the same individuals. Treatment of the data has included correlations between total scores, mental growth curves, an analysis of items for increase in per cent passing from adolescence to adulthood, and an analysis of the predictive value of preschool items. This latter item analysis was based on: correlations between Stanford-Binet (Forms L and M) preschool items and both adolescent Stanford-Binet IQ and adult WAIS verbal and performance quotients, and comparisons between scores on special scales constructed from the several kinds of preschool items and later scores obtained on comparable adult WAIS scales. Because for one subject the WAIS was incomplete and educational attainment was not ascertained, and for another the 1931 Stanford-Binet IQ was invalid, the number of total subjects on which any one analysis is based varies from 109 to 111 depending upon which variables are involved in the analysis. No attempt was made in the two follow-ups to continue the procedure used in the initial examination of testing children within one month of a birthday; a larger deviation in the age distribution was a consequence.

Table 1: *Correlations between test scores at the several administrations*

Test	1941 S-B (L)	1956 S-B (L)	1956 WAIS Full	1956 WAIS Verbal	1956 WAIS Performance
1931 S-B (L and M)	.65	.59	.64	.60	.54
1941 S-B (L)		.85	.80	.81	.51
1956 S-B (L)			.83	.89	.46
1956 WAIS Full				.87	.84
1956 WAIS Verbal					.59

RESULTS AND DISCUSSION

Constancy of the IQ

A finding of primary interest is that an individual's measured intelligence in relation to others of his same age is demonstrated to be more stable than previous studies suggest. These results are presented in Table 1. It will be noted that the correlation between preschool (1931) intelligence, and adult (1956) intelligence as measured by the full scale WAIS is .64. The measure of preschool intelligence was a combination of the two forms of the Stanford-Binet which became the 1937 Revision, and the measure of adult intelligence was a combination of the verbal and performance scales of the WAIS; thus the indices of intelligence at both ages were based on the most adequate measurements available when the data were collected. The reliability of each measurement was sufficiently high so that correcting for unreliability increases the correlation only slightly: from .64 to .69. The degree of correspondence over a 25-year period is not significantly different from the correlation of .65 (uncorrected) found between preschool and adolescent Stanford-Binet IQs after an interval of 10 years.

Also of note from Table 1 is the correlation of .80 between adolescent (1941) Stanford-Binet and adult (1956) WAIS. Although this is somewhat lower than the correlation of .85 between the adolescent and adult administrations of the Stanford-Binet, neither correlation is significantly different from the correlation of .83 between Stanford-Binet and WAIS administered to our subjects on the same day when they were adults. Apparently whatever the Stanford-Binet measures at preschool and at adolescence is still being measured by the WAIS at adulthood.

A comparison of the columns of Table 1 shows a close similarity, both in absolute amount and in trend, between the 1956 verbal WAIS and the 1956 full scale WAIS when each is correlated with the Stanford-Binet administered at the three age levels. The performance WAIS correlations, on the other hand, are both lower in value and opposite in trend. The reason for this reversal in trend needs further study to determine whether the results reflect an increase in specialization of intelligence with age or are due to an increase in verbal content of the Stanford-Binet at the higher levels.

Growth and Decline of Intellectual Abilities

Our data contribute to the accumulating evidence that intellectual potential does not cease to increase at 16 years in contradiction to the conclusions of Terman and others. Table 2 summarizes our data for comparison of obtained IQs. In view of the conformance between the 1931 and 1941 Stanford-Binet IQs and between the standard deviations at the three ages, the fact that the 1956 Stanford-Binet IQ exceeds the 1941 Stanford-Binet IQ by a statistically significant amount is interpreted as a consequence of intellectual growth in these subjects beyond 16 years. The increase of 11 points in IQ corresponds to a gain in mental age score of 20 months beyond expectancy. The finding of a continuation of mental growth into adulthood is consistent with the findings of Bayley (1955) and others (Bayley & Oden, 1955). The implications of the difference between present and previously held

Table 2: *Means and standard deviations of the two tests at the three administrations*

Test	1931		1941		1956	
	M	*SD*	*M*	*SD*	*M*	*SD*
Chronological age	4.0	1.1	13.6	1.4	29.5	1.5
S-B IQ	112.8^a	15.9^a	112.3^b	16.4^b	123.6^b	15.0^b
WAIS IQ						
Full					108.9	11.0
Verbal					110.4	13.3
Performance					105.0	11.5

[a] *Composite Form L and Form M.*
[b] *Form L.*

Table 3: *Stanford-Binet scores for adults by year of age*

1956 Chronological age	*N*	S-B Mean 1956 IQ	S-B Mean 1956 MA
27.0–27.9	16	121.3	18.2
28.0–28.9	25	125.5	18.9
29.0–29.9	23	126.8	19.0
30.0–30.9	20	123.5	18.5
31.0–31.9	24	121.0	18.1

Table 4: *WAIS scores for adults by year of age*

1956 Chronological age	Mean 1956 WAIS standard scores			
	N	Full	Verbal	Performance
27.0–27.9	16	119	67	51
28.0–28.9	24	127	72	56
29.0–29.9	23	129	73	55
30.0–30.9	20	126	72	54
31.0–31.9	24	119	70	50

conclusions regarding mental growth into adulthood deserve consideration. Bayley (1955, p. 817) suggests alternative explanations including the new generations' having more opportunity to develop their intellectual capacities than did their parents. Inhelder and Piaget (1958, p. 337) have pointed out the probability that Greek children were behind our own in growth of logical thinking and suggest that the age at which formal thinking begins "may be a product of progressive acceleration of individual development under the influence of education."

The method of computing WAIS IQs assumes continuation of intellectual growth up to 34 years. The mean WAIS IQ of our group is lower in numerical value than expectancy based on the mean 1941 Stanford-Binet IQ. It will be noted, however, that the WAIS mean IQ is approximately one *SD* above 100, whereas the 1941 Stanford-Binet mean IQ is less than one *SD* above 100; so that in terms of relative deviations, the WAIS mean IQ is higher than the Stanford-Binet IQ, although not to a statistically significant degree. The verbal half of the WAIS approaches the 1941 Stanford-Binet results more closely than does the full scale, and the performance half less so.

The addition of the present data to those previously collected for this group of subjects gives us a third segment of the mental growth curve. Thus we can plot mental age against chronological age for the years 2 through 5, the years 11 through 15, and now the years 27 through 31. Two subjects were younger and one was older than these years and are omitted from this section. Table 3 summarizes the current data for the Stanford-Binet. . . .

Before we interpret what this lack of change during this 5-year period may mean, we must consider the possible ceiling effect. The 1937 Stanford-Binet has insufficient top for adults and therefore this flattening of the curve might be an artifact. To examine this possibility, a similar analysis was made of the results on the WAIS for these subjects, inasmuch as a low ceiling is not a limitation of that test. One of the subjects had a partial WAIS only and is omitted for this reason. . . . Table 4 summarizes the data for the WAIS. Analysis of variance for these three scores again shows the variation with chronological age not to approach significance. . . .

Further study to determine the role of ceiling effect was made by analyzing only those cases showing no or few passes at the Superior Adult III level of the Stanford-Binet. Also, because our group of subjects is somewhat above average, a separate study was made of those cases clustering around IQ 100. These additional studies, however, served only to substantiate the results just reported. The point at which mental growth as measured by these instruments terminates and at which the curve assumes a downward direction is not established by these data.

Sex Differences

Tests such as the Stanford-Binet are intentionally constructed to eliminate or to minimize sex differences. Consequently most investigators, including the present ones in earlier studies, have given little attention to this variable. It is in this area, however, that some of the most cogent findings of the present study emerge. As shown in Table 5, on all three administrations of the Stanford-Binet, female subjects achieved a higher mean score than did males. On the first two administrations, this difference was significant beyond the .01 level. On the 1956 testing, it did not reach significance. The males increased their scores from 1941 to 1956 to a greater extent than did the women. For the males, this increase was 13.7 and for the females it was 9.6; the difference of 4.1 is significant beyond the .01 level. Both females and males show a negative correlation between the 1941 Stanford-Binet IQ and the gain in IQ from 1941 to 1956, but the one for females is larger. For men this correlation is −.23 and for women −.67; the difference between these correlations is significant beyond the .01 level. Dividing the male and female groups according to whether they stopped their education at high school graduation or went beyond does not change the findings. . . . The decrease in standard deviation for females from 1941 to 1956 as shown in Table 5 reflects that these findings are due to the tendency of women with IQs which are high at adolescence not to increase their scores by adulthood as much as does the rest of the group.

A study of the individuals scoring high (from 130 to 150 IQ) on the 1956 Stanford-Binet yields further information. Females in this group had always scored high relative to the group: mean IQ of 125.9 in 1931, 127.8 in 1941, and 137.5 in 1956. For the males, however, becoming a member of this group represented a gain from an earlier lower point with most of the gain achieved by adolescence: mean IQ of 115.9 in 1931, 126.2 in 1941, and 140.7 in 1956. (It must be remembered that these 1956 Stanford-Binet IQs are, on the average, 11 points high, since they do not include a correction for the fact that there was a mean IQ increase of 11 points for the total sample.) Slightly but not significantly more women than men in this subgroup stopped education with high school graduation. Of those who started college, not a single man failed to finish, whereas one-third of the women dropped out before finishing. This leaves us with the familiar chicken-egg dilemma. Did the women drop out of college because their relative brightness as compared with men decreased—their promise not being realized in the same way as was that of the men? Or did the earlier termination of their academic education make the intelligence tests administered to them as adults relatively more difficult for them?

Table 5: *Sex differences on the Stanford-Binet*

Test	Male	Female	Difference
1931 L-M			
N	51	59	
M	108.5	115.8	−7.3
SD	14.4	13.3	
*CR**			2.79
1941 L			
N	52	59	
M	107.9	116.1	−8.2
SD	19.0	13.7	
*CR**			2.58
1956 L			
N	52	59	
M	121.6	125.7	−4.1
SD	19.6	10.6	
*CR**			1.36

** A value of 2.57 is significant beyond the .01 level and a value of 1.96 beyond the .05 level.*

An analysis of the relation between IQ at adolescence and continuation of education for the total group was made. There is no significant difference for either sex between mean 1941 IQ of those who stopped their education with graduation from high school and the mean 1941 IQ of those who attended college but did not graduate (males: 100.7 and 102.5; females: 113.8 and 113.0). The mean 1941 IQ, however, of both the 17 males and the 14 females who graduated from college is significantly higher (124.6 and 125.0, respectively) than for those who did not graduate from college. Gain in IQ, on the other hand, is not consistently related to educational attainment. Mean IQ gains for males who did not go beyond high school graduation, who entered college but did not graduate, and who graduated from college were 14.7, 11.7, and 11.8, respectively. Corresponding mean gains for females were 8.2, 12.3, and 9.3, respectively. The differences between the mean gains for either sex are not large enough to be statistically significant with the small number of subjects involved (males: 22, 13, and 17 for successive educational attainments; females: 22, 22, and 14).

Because other investigators of adult intelligence have not reported sex differences of the nature found here, we cannot look beyond our own data for the answers to some of the questions which are posed. This seems to open up an area worthy of further study designed explicitly to explore this problem.

Increases in Abilities from Adolescence to Adulthood

With the establishment of the fact that there is increase in intelligence scores between adolescence and adulthood, one becomes interested in identifying the kinds of abilities which increase most. Within the range of abilities covered by the adult levels of the Stanford-Binet, it was possible to shed light on this question by comparing the proportion of our group passing each item as adolescents and as adults. The percentages of the present sample passing each item were calculated for the 1941 and 1956 results. Items occurring below a subject's basal age were assumed to have been passed; those occurring above a ceiling were assumed to have been failed. Since nearly all subjects passed in 1956 all those items through Level 14, it was possible to make meaningful comparisons for the four adult levels only. Table 6 lists each of the 26 items at these levels, arranged according to the magnitude of the increase in per cent passing over the 15-year period. The final column in the table classifies items by "type," following a system suggested by Davis (1941), into rote memory (M), reasoning (R), and vocabulary (V). The reasoning subscale has been further divided by Robinson into those items which appear to involve mental manipulation of practical, concrete concepts (RP) versus those dealing with more abstract concepts (RA). One might wonder whether those items having the lowest per cent passing in 1941 would by artifact show the most increase to 1956. The correlation between per cent passing in 1941 and increase from 1941 to 1956 was found, however, to be only .06.

Table 7 presents a comparison of the data which demonstrates the differential increases made by our subjects in the four classifications of items. Immediate rote memory items and those

Table 6: *Items of Stanford-Binet (L) adult levels by magnitude of change in per cent passing from adolescence to adulthood*

Rank	Location	Item	Per cent Passing[a]		Increase	Chi square[b]	Type
			1941	1956			
1	AA-5	Proverbs I	22.5	81.1	58.6	76.26	RA
2	SAI-1	Vocabulary	35.1	92.8	57.7	80.07	V
3	SAII-1	Vocabulary	13.5	67.6	54.1	67.28	V
4	SAI-6	Essential Similarities	31.5	82.9	51.4	52.42	RA
5	SAI-5	Sentence Building	30.6	77.5	46.9	49.04	V
6.5	AA-6	Ingenuity	30.6	73.9	43.3	38.01	RP
6.5	SAIII-1	Vocabulary	1.8	45.0	43.2	52.84	V
8	AA-3	Difference between Abstract Words	49.5	91.9	42.4	48.05	V
9	SAIII-3	Opposite Analogies	8.1	48.7	40.6	44.89	V
10.5	AA-2	Codes	37.8	77.5	39.7	35.73	RP
10.5	SAII-5	Reconciliation of Opposites	29.7	69.4	39.7	34.89	RA
12	AA-1	Vocabulary	59.4	98.2	38.8	49.41	V
13	SAII-4	Proverbs II	6.3	44.1	37.8	42.81	RA
14	AA-7	Memory for Sentences V	36.9	72.1	35.2	27.63	M
15	SAII-2	Finding Reasons II	14.4	47.8	33.4	28.68	RA
16	AA-8	Reconciliation of Opposites	54.9	86.5	31.6	26.64	RA
17	AA-4	Arithmetic Reasoning	41.4	72.1	30.7	21.22	RP
18	SAIII-2	Orientation: Directions II	14.4	41.4	27.0	20.15	RP
19	SAI-3	Minkus Completion	24.3	50.5	26.2	16.18	V
20	SAI-4	Repeating 6 Digits Reversed	29.7	54.1	24.4	13.45	M
21	SAIII-5	Reasoning	3.6	24.3	20.7	19.82	RP
22	SAII-3	Repeating 8 Digits	17.1	35.1	18.0	9.34	M
23	SAIII-6	Repeating 9 Digits	3.6	16.2	12.6	9.89	M
24	SAIII-4	Paper Cutting II	11.7	22.5	10.8	3.57	RP
25	SAI-2	Enclosed Box Problem	50.4	55.9	5.5	0.62	RP
26	SAII-6	Repeating Thought of Passage	24.3	27.9	3.6	0.09	M

[a] *N = 111.*

[b] *A chi-square value of 10.83 is significant beyond the .001 level, 6.64 beyond the .01 level, and 3.84 at the .05 level. Hence all but the last three items of this table show a significant increase in per cent passing from adolescence to adulthood.*

involving practical reasoning showed smaller increases than those involving abstract reasoning and vocabulary. This finding tends to confirm the conclusion of most investigators that growth of functions dependent on continued learning and on the building of concepts is more likely to continue into adulthood than is growth of memory or performance functions.

These data make it clear that one should not present conclusions about the growth of intelligence into the adult years without specifying what kinds of intelligence one means and how one is measuring these. If one administered an intelligence test consisting solely of items measuring vocabulary and abstract reasoning, he would find more increase of "intelligence" into adulthood than if he used a test heavily weighted with memory and practical reasoning items. Thus one cannot realistically expect ever to describe the shape of *the* mental growth curve.

A finding in these data which has important implications when it comes to trying to predict amount of subsequent mental growth in an individual adolescent is the negative relationship between level of ability at adolescence and gain in mental age score from adolescense to adulthood. We have found, for example, that the higher the 1941 Stanford-Binet score, whether expressed as IQ or MA, the slightly lower the rate of increase (MA increase/CA increase) to 1956 for the group of 45 subjects who were 14 to 16 years of age in 1941. (For this analysis, subjects less than 14 years of age in 1941 were eliminated in order to minimize the operation of differential rates of mental growth in adolescence.) The correlations expressing this are not significantly different from zero. Comparable correlations for males and females considered separately are consistent with the findings on sex differences discussed earlier in that the females showed a significant negative relationship, whereas the males did not. . . .

Thus, although the adult scores on intelligence tests can be predicted with some accuracy from the score one obtains as an adolescent (as indicated by the correlations in Table 1), we have not isolated factors which permit us to predict the amount of mental age gain. The fact that adult intelligence can be predicted from adolescent intelligence with such a high degree of accuracy despite the lack of relation between adolescent intelligence and amount of gain to adulthood is, of course, due to the fact that the amount of gain at these levels represents such a relatively small proportion of the total "amount" of intelligence. We can anticipate that there will be an increase in mental age from adolescence to adulthood, but we cannot predict which subjects will gain most. We do know that amount of gain is not dependent upon level of initial mental age, on the extent of education, or on the level of ancestral intelligence (Bradway & Robinson, 1961), in contrast to mental growth in the earlier years where these factors are important (Bradway, 1945a). Our only positive finding in this area is the one pointed out earlier, which is that adolescent girls with high IQs can be expected to increase less in mental age than will girls who test lower, or than males who test at any level.

Table 7: *Comparison of four subscales on increase from adolescence to adulthood*

Scale	Number of items	Rank order among 26 items		Mean of increases in per cent passing
		M	Mdn.	
Vocabulary	8	8.1	7.3	43.7
Reasoning				
Abstract	6	9.9	11.8	42.2
Practical	7	17.4	18.0	25.1
Memory	5	21.0	22.0	18.7

Prediction of Adult Intelligence

The next questions which arise are those of how early can adult intelligence be predicted, and what are the best measures for predicting it. Or, what are the origins of adult verbal and nonverbal abilities?

An analysis of the predictive value of Stanford-Binet preschool items to adult WAIS scores can help answer these questions. At the time of the 1941 study, the principal investigator devised four subscales from the preschool Form I, and Form M Stanford-Binets and computed four subscale quotients for each subject (Bradway, 1945b). These were labeled "verbal quotient," "nonverbal quotient," "memory quotient," and "number concept quotient." Because the Stanford-Binet at the adolescent and adult levels is an almost wholly verbal instrument, it was not possible in that earlier study to determine whether or not one maintains his relative superiority and inferiority in these categories over any length of time. The administration of the Wechsler Adult Intelligence Scale in the present study, however, allows us to make such "profile" comparisons.

Corresponding subscale indices were derived from the WAIS by using Comprehension, Similarities, Information, and Vocabulary for comparison with the 1931 verbal quotient; the Performance scale for comparison with the nonverbal quotient; Digit Span with the memory quotient; and Arithmetic with the number concept quotient. Inasmuch as the WAIS results are in standard scores, all that was necessary to obtain indices for these four WAIS categories which were comparable with each other was to divide the verbal total by four and the performance total by five. The mean of these four indices was then obtained for each subject, and his score for each category was expressed as a deviation from his own mean. A similar method was used for the 1931 scores, except that each of the four subscale scores had to be converted into standard scores before the individual profiles of deviation scores could be figured.

The stability of pattern over a period of 25 years was tested by the correlation between the deviation scores calculated as each individual's deviation from his own mean for the 1931 subscales, and the deviation scores also calculated as each individual's deviation from his own mean for the 1956 WAIS. Three of these four correlations are significant in the positive direction. For the verbal component, the Pearsonian r is .236 significant at the .02 level; for the nonverbal component, .191 significant at the .05 level; and for the memory component, .301 significant at the .01 level. For the number concept component, (figured on 79 subjects rather than 110 because for 31 subjects no such ratio could be computed at the preschool ages) the correlation was only .089, not significant, suggesting that relative strength or weakness in this ability is not stable over a 25-year period. Inasmuch, however, as one's score on this scale contributes to his deviation scores on the other three, further analysis is necessary before drawing such a conclusion.

To study the predictive value of preschool items in more detail, biserial correlations of each Stanford-Binet preschool item to adult WAIS verbal quotient and adult WAIS performance quotient were calculated. This analysis is summarized in Tables 8 and 9 in which are listed

Table 8: *Correlation of predictive preschool items of forms L and M of the Stanford-Binet with WAIS verbal quotients*

Location	Item	r_{bis}	Groups by years of age[a]	Type[b]
M II ½ 6[c]	Obeying Simple Commands	.93***	2½	V
L III 2[c]	Picture Vocabulary	.77**	2½	V
L III 4[c]	Picture Memories	.85***	2½	M
M III 3	Objects by Use	.66*	2½	V
M III 6	Repeating Three Digits	.67*	2½	M
L III ½ 2	Picture Vocabulary	.84**	4	V
L III ½ 5	Objects by Use	.98***	2½	V
L III ½ 6[c]	Comprehension I	.86**	4	V
M III ½ 6	Comprehension I	.74**	2½	V
L IV 1[c]	Picture Vocabulary	.85***	4	V
L IV 2[c]	Objects from Memory	.79**	4	M
L IV 4[c]	Pictorial Identification	.80**	4	V
L IV A[c]	Memory for Sentences I	.80**	4	M
M IV 6	Memory for Sentences I	.70*	2½	M
L IV ½ 6	Opposite Analogies I	.73**	4	V
M IV ½ 1	Animal Pictures	.64*	4	NV
M IV ½ 5	Materials	.74**	4	V
L V 5	Memory for Sentences II	.64*	4	M
M V 2	Number Concept of Three	.74**	4	NC
M V 3[c]	Pictorial Similarities and Differences	.61*	5½	V
L VI 4	Number Concepts	.71**	5½	NC
L VI 5	Pictorial Likenesses and Differences	.72***	5½	V
M VI 2	Copying Bead Chain	.55*	5½	NV
M VI 3[c]	Differences	.60*	5½	V
M VII 4[c]	Three Digits Reversed	.62*	5½	M

[a] *There were 17 subjects in the 2½-year-old group, 16 in the 4-year-old, 17 in the 4½-year-old, and 21 in the 5½-year-old. The other groups (2-year-olds, 3-year-olds, 3½-year-olds, and 5-year-olds) were not large enough for valid calculations.*

[b] *These refer to the Bradway subscales described previously (Bradway, 1945b).*

[c] *Indicates used in 1960 L-M Revision of the Stanford-Binet.*

* *Significant beyond .05 level.*

** *Significant beyond .02 level.*

*** *Significant beyond .01 level.*

those items whose biserial *r*'s were significant beyond the .05 level for the indicated age groups.

The figures in the column headed "Groups by years of age" indicate the age group for which the corresponding biserial *r* was figured. Since there was no way of combining the age groups, biserial *r*'s were figured for each 1931 age group for which an item was pertinent, which was defined for purposes of this analysis as each age group in which the number of subjects was 15 or more and for which the per cent passing the item was between 25 and 75. The distribution by type of the total 79 items for which biserial *r*'s at one or more age levels were computed were as follows: 37 verbal, 24 nonverbal, 12 memory, and 6 number concept. The type designation follows that previously described by Bradway (1945b). Comparison of the frequency of each type found to have significant predictive value at one or more age levels and the total number of such items studied shows the following: of 37 verbal items, 14 (38%) predicted WAIS verbal IQ and 14 (38%) predicted WAIS performance IQ; of the 24 nonverbal items, 2 (8%) predicted WAIS verbal IQ and 5 (21%) predicted WAIS performance IQ; of the 12 memory items, 7 (59%) predicted WAIS verbal IQ and 5 (41%) predicted WAIS performance IQ; of the 6 number concept items, 2 (33%) predicted WAIS verbal IQ and 3 (50%) predicted WAIS performance IQ. This kind of analysis is only suggestive because the number of items is small, but it supplements in more detail the findings for the subscales reported above. In summary: verbal items predict both verbal and performance adult ability better than do nonverbal items; each kind of memory item (memory for digits forward and reversed, for sentences, for objects, for pictures, and for bead chain) predicts both WAIS verbal and performance ability; number concept items are incon-

Table 9: *Correlations of predictive preschool items of forms L and M of the Stanford-Binet with WAIS performance quotients*

Location	Item	r_{bis}	Groups by years of age[a]	Type[b]
M III 2[c]	Picture Vocabulary	.90***	2½	V
L III ½ 6[c]	Comprehension	.75*	4	V
L IV 1[c]	Picture Vocabulary	.75*	4	V
L IV 2[c]	Objects from Memory	.71*	4	M
L IV 4[c]	Pictorial Identification	.87***	4	V
L IV 5[c]	Discrimination of Forms	.71*	4	NV
M IV 3[c]	Opposite Analogies I	.70*	4	V
M IV 4[c]	Pictorial Identification	.70*	2½	V
M IV 6	Memory for Sentences I	.75*	4	M
L IV ½ 4	Materials	.68*	4½	V
L IV ½ A	Pictorial Identification	.71*	4	V
M IV ½ 1	Animal Pictures	.63*	4	NV
M IV ½ 3	Repeating Four Digits	.70*	2½	M
M IV ½ 5[c]	Materials	.69*	4	V
L V 4[c]	Copying a Square	.74*	4	NV
L V 6	Counting Four Objects	.62*	4½	NC
M V 2	Number Concept of Three	.75**	4	NC
M V 5	Comprehension II	.78***	5½	V
L VI 1	Vocabulary	.86***	5½	V
L VI 2	Bead Chain from Memory	.68* .58*	4½ 5½	M
L VI 3[c]	Mutilated Pictures	.82** .77***	4½ 5½	NV
L VI 5	Pictorial Likenesses and Differences	.53*	5½	V
M VI 2	Copying Bead Chain	.94***	5½	NV
M VI 5	Counting 13 Pennies	.67**	5½	NC
M VI 6	Opposite Analogies I	.73**	5½	V
L VII 1[c]	Picture Absurdities I	.61*	5½	V
M VII 4	Three Digits Reversed	.60*	5½	M

[a] There were 17 subjects in the 2½-year-old group, 16 in the 4-year-old, 17 in the 4½-year-old, and 21 in the 5½-year-old. The other groups (2-year-olds, 3-year-olds, 3½-year-olds, and 5-year-olds) were not large enough for valid calculations.

[b] These refer to the Bradway subscales described previously (Bradway, 1945b).

[c] Indicates used in 1960 L-M Revision of the Stanford-Binet.

* Significant beyond .05 level.

** Significant beyond .02 level.

*** Significant beyond .01 level.

sistent in their predictive value—the only significant negative correlation obtained being for a number concept item.

When one compares these results with those for the subscale deviation scores reported above, similarities are found. Both analyses suggest the greater value of verbal and memory items over nonverbal items in predicting adult abilities, and the unreliability of number concept items at the preschool level as predictors of adult abilities.

Measurement of Intelligence

After reviewing these data we are in a better position to approach the problems of what index or indices to use in expressing adult intelligence and how best to predict it. If the 1937 Stanford-Binet is used to measure adult intelligence, a use for which it was not specifically designed, it is apparent that a correctional factor of some kind must be applied to IQs obtained according to the published instructions. Merrill kindly applied the deviation scores provided in the 1960 Revision (Terman & Merrill, 1960) to our data. The resulting intelligence quotients are 109.3 in 1931, 109.4 in 1941, and 119.4 in 1956. These findings of identity between preschool and adolescence and a 10 point rise to adulthood indicate that the correction inherent in the 1960 deviation scores does not solve the problem. For our research purposes we rejected the use of a multiplicative factor because this would have changed the standard deviations, and used instead a subtractive factor which, since the standard deviations were nearly identical, was sufficient to equate the adult set of scores with the two previous ones. This, although a satisfactory solution for research purposes, may not be the optimal solution from the point of view of the clinician. Perhaps the possibility of revised deviation tables, accounting for greater increase into adulthood than do those published with the 1960 Revision, should be explored.

This problem is not encountered in measuring adult intelligence with the WAIS which was specifically designed for the measurement of adult intelligence. Moreover, our data demonstrate that there are different kinds of ability which have a degree of constancy throughout life. The structure of the WAIS is a recognition of this.

When new tests are constructed, those for adults would do well to provide valid measurements of different kinds of abilities, because the practicing clinician needs such information to evaluate selective impairment of intellectual functioning and for vocational guidance and occupational placement. In contrast, a preschool test heavily weighted with verbal items should be a more valid predictor of adult abilities, both verbal and nonverbal, than one utilizing many performance-type items. Moreover, the former would be consistent with the usual purpose for which preschool tests are used in that the prediction of school performance puts a heavy demand on verbal abilities.

The present data demonstrate that the defining of adult intelligence is complex. It is not expressed by a single index in terms of abilities, in terms of one's place relative to others of the same age group, or of one's location on a scale of development and decline. As Cameron and Magaret (1951, p. 158) have said in alluding to the use of tests in defining behavior, "Performance on test items alone can no more define biosocial immaturity than can a refractive index define the complex process of seeing."

SUMMARY

A group of 111 subjects who had been tested with the Stanford-Binet (S-B) at the preschool and adolescent levels were administered the S-B and WAIS at adulthood (25 years after initial testing). Correlations of preschool IQs with adult S-B and full WAIS IQs are .59 and .64. Correlations of adolescent IQs with adult S-B and full WAIS IQs are .85 and .80. The S-B IQ shows a mean increase of 11 points from adolescence to adulthood, which corresponds to a mental age gain of 20 months beyond what would be expected if mental growth ceases at 16 years. The data were analyzed in an attempt to determine the point at which mental growth, as measured by these tests, ceases, but it could not be determined.

The males show significantly more IQ gain from adolescence to adulthood than do the females ($p < .01$). Factors associated with amount of gain were not isolated except that girls with high IQs at adolescence increase less than do girls who initially test lower or than males who test at any level. There is no evidence that educational attainment is related to amount of gain.

Analysis of increases in per cent passing S-B items from adolescence to adulthood indicates that intellectual growth after adolescence is greater in abstract reasoning and vocabulary than in rote memory and practical reasoning.

Analysis of the value of S-B preschool items in predicting adult WAIS scores shows some stability of pattern of individual differences in verbal versus nonverbal versus memory superiority over the 25 years. It also shows the greater value of verbal and memory items over nonverbal items in predicting both verbal and performance adult IQs, and the unreliability of number concept items as predictors of adult abilities.

The data are interpreted as demonstrating that adult intelligence is complex and cannot be meaningfully expressed by a single index.

REFERENCES

BAYLEY, NANCY. On the growth of intelligence. *Amer. Psychologist*, 1955, *10*, 805–818.

BAYLEY, NANCY, & ODEN, MELITA H. The maintenance of intellectual ability in gifted adults. *J. Gerontol.*, 1955, *10*, 91–107.

BRADWAY, KATHERINE P. An experimental study of factors associated with Stanford-Binet IQ changes from the preschool to the junior high school. *J. genet. Psychol.*, 1945, *66*, 107–128. (a)

BRADWAY, KATHERINE P. Predictive value of Stanford-Binet preschool items. *J. educ. Psychol.*, 1945, *36*, 1–16. (b)

BRADWAY, KATHERINE P., & ROBINSON, NANCY M. Significant IQ changes in twenty-five years: A follow-up. *J. educ. Psychol.*, 1961, *52*, 74–79.

CAMERON, N., & MAGARET, ANN. *Behavior pathology*. Boston: Houghton Mifflin, 1951.

DAVIS, F. B. The derivation of three subscores from the 1937 Revision of the Stanford-Binet Scales. *J. consult. Psychol.*, 1941, *5*, 287–291.

INHELDER, B., & PIAGET, J. *The growth of logical thinking from childhood to adolescence*. New York: Basic Books, 1958.

TERMAN, L. M., & MERRILL, MAUD A. *Stanford-Binet Intelligence Scale*. Boston: Houghton Mifflin, 1960.

CONCEPT DEVELOPMENT

59

A Study of Children's Moral Judgments

Ronald C. Johnson

Johnson's study is an attempt to test Piaget's ideas of the development of morality. The test items bear a relationship not only to Piaget's questions but also to the Anderson Incomplete Stories (45). Johnson's approach is cross-sectional and includes age, intellectual, socioeconomic, and parental attitude influences as variables. His study demonstrates the effects of intelligence and environment upon the stated moral judgment of children and adolescents.

The development of moral judgment in the child has been of considerable interest to the social scientist. Among the students of human development, Jean Piaget has been, perhaps, most influential in this area of study (1932). His studies of the child's developing understanding of causality, of physical and temporal reality, and of ethics and moral judgment have been extremely fruitful. This study proposes to test some of Piaget's ideas concerning developmental changes in moral judgment.

Moral judgment, for Piaget, consists of a number of areas. These include: (a) immanent justice—the belief in the existence of automatic punishments which emanate from things themselves; (b) moral realism—a belief that acts should be judged in terms of consequences, not on the basis of the motive behind the act; (c) belief that punishment should be retributive vs. belief that punishment should be restitutive (merely restore the equilibrium destroyed by the pun-

---

From *Child Development*, 1962, *33*: 327–354. Reprinted with permission of the Society for Research in Child Development.

ished act); (d) acceptance or rejection of the idea that the more severe punishment is more efficacious; and (e) choice of collective (essentially, guilt by association) or of individual responsibility for punishable acts.

. . . Piaget sees three forces interacting to produce developmental change in moral judgment: adult constraint, peer group cooperation and reciprocity, and the changing character of the child's mind. He says, ". . . We have three processes to consider: the spontaneous and unconscious egocentrism belonging to the individual as such, adult constraint, and cooperation." ". . . cooperation alone can shake the child out of its initial state of unconscious egocentrism; whereas constraint acts quite differently and strengthens egocentric features of moral realism until such time as cooperation delivers the child both from egocentrism and from the results of this constraint" (1932, p. 184). The interaction of these factors, the changing mind of the child, the amount of adult constraint, and the amount of peer group cooperation and reciprocity, to Piaget, cause developmental differences in systems of responsibility and type of moral judgment used by children of different ages.

Piaget and his assistants studied a sample of about 100 children aged about 6 to 12, from homes of lower socioeconomic groups, in Geneva, Switzerland. They were questioned and interviewed individually.

So far as this writer can determine, Piaget's study was conducted primarily to show that developmental changes occur in moral judgment and then to determine how the factors of adult constraint, peer group cooperation, and qualitative changes in thought processes interact to produce these changes in moral judgment.

Piaget shows that age changes in belief do occur. Younger children accept the concepts of immanent justice and moral realism, believe that retribution should be the primary basis for punishment, and believe in the efficacy of severe punishment. Older children reject the concepts of immanent justice and moral realism and believe that less severe, restitutive punishments are more appropriate. The youngest and older segments of his sample accepted communicable responsibility as just, while a middle age group did not. This writer was unable to find any statement by Piaget that these changes necessarily occur at a specific age, even in his own sample. Piaget seems most aware that changes in belief would occur at different times in other segments of society, (e.g., see his statement concerning immanent justice [1932, pp. 255–256]), though he would say that the direction of change would be the same.

It does not seem that he has achieved his second objective, that of showing how the interaction of constraint, cooperation, and mental change cause this change to occur. While the interview technique that Piaget used may have decided advantages, it would seem impossible, through interviews, to determine the relative significance of three factors, each acting on moral judgment and each acting upon the other two factors. None of the material presented seems, to this writer, to be able to separate out the influence of each of these three factors in terms of anything but Piaget's own subjective judgment. Therefore, it would seem that Piaget's purposes are only half fulfilled; he has shown that age change in moral judgment does occur, but has not shown exactly what influence various factors have in producing the change. . . .

It would seem, from a review of the literature, that Piaget's discovery of an age change in moral judgment has been fully confirmed, though the specific age of the change would depend on the nature of the sample and the level of difficulty of the questions asked. The question of how adult constraint, peer group cooperation, and qualitative changes on mutual structure interact to produce this change (or if they do at all) seems to be largely unresolved. A third question, that of whether there is such a thing as moral judgment, as such, or whether moral judgment merely consists of a number of specific areas of response that are essentially unrelated to one another seems an even more basic question that has not yet been answered at all satisfactorily.

This study was concerned first with an attempt to determine the degree of interrelation within and between areas of moral judgment. This study dealt secondly with the relation of various antecedent conditions to moral judgment. There was an attempt to discover the relation of adult constraint, "egocentricity," age, sex, IQ, and parental occupational level to moral judgment. Though certainly one should also measure peer group cooperation, this seemed impossible in a sample sufficiently large to give a reliable determination for the primary problem. Previous findings concerning the effect of peer solidarity are uniformly negative, indicating that, of the variables, this might be omitted.

This study, using a different method of gaining information, different questions, and different age groups cannot be considered a test of the validity of Piaget's own work, since it was not directly comparable. However, it did attempt to test Piaget's ideas.

METHOD

Sample

This study included all the children (all of those present on the day when their classes were tested) in grades 5, 7, 9, and 11 in a midwestern public school system. The distribution of the sample by grade and by sex was as follows: grade 5, 93 boys, 97 girls; grade 7, 97 boys, 76 girls; grade 9, 125 boys, 143 girls; and grade 11, 84 boys and 92 girls. The total sample consisted of 807 subjects.

Data gathered from this sample will be discussed in terms of grade, not age, although age was the significant variable in Piaget's own analysis. This school district, like many others, follows a policy of passing almost 100 per cent of its students. An actual count of retarded and of advanced students at grades 9 and 11 showed that only 4 per cent were retarded and less than 1 per cent advanced in grade in comparison with their agemates. An analysis of grade, therefore, is actually (except for this small proportion of retarded and advanced students) an analysis of age as well.

A comparison of the occupational level of the parents of these children with that of the urban population of the United States indicates that the sample was a very close approximation of the general population, as measured by the Minnesota Occupational Scale (1950). Information concerning IQ was available in 732 cases of the total sample of 807. The mean IQ of the sample was 105.9. The sample closely approximated U. S. urban norms in intelligence as well as in occupational level.

Measuring Devices

The major purpose of this study was to determine the degree of interrelation between responses to questions formulated to represent different areas or kinds of moral judgment. The writer's initial problem was that of devising a test of moral judgment within which a number of questions could be found having to do with each of the areas under discussion. A number of stories were made up, modeled quite closely after those of Piaget. The stories illustrated five types of moral judgment: immanent justice, moral realism, retribution and expiation vs. restitution and reciprocity, the efficacy of severe punishment, and communicable responsibility.

Piaget's questions were generally quite simple, since they were designed for use with 6- to 12-year-olds. The writer wished to use an older sample; so that the moral judgment test could be administered as a paper and pencil test and so that a test of "abstractness-concreteness" might be used. For these reasons, a number of new questions, based on those used by Piaget, were devised. The work of Dennis (1953) in the field of animism, another area studied by Piaget, clearly indicates that the level of questions asked affects the level of response more than any qualitative age change in approaches to problem solving. The assumption was therefore made (and later confirmed) that with the proper questions there would be considerable variation in response even in a much older group than Piaget's.

A number of question stories similar to Piaget's were discussed with several people familiar with Piaget's work to find those stories most similar to Piaget's own. Since Medinnus (1957) believed that differences in the "level of abstraction" in the stories might account for differences in the type of response obtained, items were eliminated and revised until they seemed of approximate equal difficulty to this writer and to Medinnus, who helped in their selection. Twenty items were retained as those likely to be intrinsically interesting to the subjects, most similar in meaning to those of Piaget, and of an equal level of difficulty.

The moral judgment test items appear below, followed by a discussion of the way that responses would, according to Piaget's results, be scored in terms of maturity.

1. Four boys went downtown one day and were looking around in the store windows. One boy said, "Let's go into this store and see if we can take something." Two of the others thought this was a good idea but the other one didn't want to. The three that wanted to told the one that didn't want to that they'd tell all the other guys that he was chicken. So he went with them. They all took a few things but then the store detective caught up with them. He caught the one that had suggested it and he caught the two that had gone along with the idea. But he didn't notice the one who hadn't wanted to take anything, even though he was right beside the others and was acting the same

as they were, so that this one didn't get caught at all.

Do you think that this boy would have been caught, too, if he had wanted to take things like the other three did?

Now why did this one get away when the other three were caught?

2. A girl named Nancy liked one of the boys in her class an awful lot, but she didn't want anyone to know it, even though she'd never been out with this boy or anything like that. She told her best girl friend about it and the girl friend promised not to tell. But then she got angry at Nancy one day and told a lot of people, so that Nancy got teased a lot, but this boy hadn't gone with her before, so she had nothing to lose on that score.

Now here is a story something like the first one:

This girl was named Sharon. She was going steady with one boy, but her parents had visitors from out of town, who had a son her age. Her parents asked her to show this boy a little of the town. She didn't like doing this because she thought it wasn't fair to her steady boy friend, but she went out with this boy to be polite. She told her best friend about this and the girl friend promised not to tell—just like in the first story. Entirely by accident, this girl friend mentioned it in front of Sharon's steady boy friend and some other kids. He got angry at Sharon because she'd gone out with someone else when they were supposed to be going steady, so he quit going with her. Sharon got teased and lost her boy friend too because of what her girl friend had said without thinking.

Now was Nancy's girl friend or was Sharon's girl friend the worst?

Why?

3. Some boys and girls were playing football in the street. A pass was thrown and the boy who ran to catch it ran up onto a yard and trampled some recently planted shrubs and flowers. What should be done to the boy: (1) make him pay for the damage done or replace the shrubs and flowers. (2) give him a licking so he won't do it again. (3) not allow him to play football.

What punishment would be fairest?

What punishment would be hardest?

4. The stores downtown are always worried about shoplifting. In the past few years the stores have been good to high school students by giving them part-time work, especially at Christmas, to provide those who don't have much money with some buying power. The police have been working on the shoplifting problem, too, arresting many more people. Shoplifting has decreased recently.

What do you think is responsible for most of the decrease, the stores helping the young people or the increased police action?

Why?

5. A group of young people were coming back to school after playing softball. Some of them had picked up rocks on the baseball diamond and were throwing them around. Finally, one boy threw a rock which broke a window in the school. The principal was nearby and heard the crash. He took them all into a room and asked them who broke the window. Some of them hadn't thrown any rocks at all, and of those who had thrown them only one broke the window. The boy who had broken the window wouldn't say that he had done it and the other boys would not tell on him.

Should the principal punish all of the boys or none of them?

Why?

6. A boy, about 16 years old, had a driver's license. His father said that he could use the car whenever he asked, as long as his parents weren't going to use it. The boy used it pretty often but one night his father told him not to, because he and the boy's mother wanted to use it as soon as the father got off work that evening. His mother was out so the boy figured that he'd use it anyway and get it back before either parent got home. He and a buddy of his took a ride around town but just as he was headed home, he got a flat. Because of the time it took fixing the flat, the boy didn't get the car home on time to beat his Dad, so that his Dad found out, and wouldn't let him use the car for a month.

Now, if he'd have been driving the car with his father's consent would he have still got the flat?

Why?

7. There was a boy named Bill whose parents were going on vacation. He had a paper route and had to get someone to take it for him for a week so that he could go with them. He got his friend, a boy named Jimmy, to take the route and Jimmy did a good job for a couple of days. Then Jim's grandfather got sick and his parents had to leave town to be with him and the boy had to go with his parents. He didn't want to go and leave Bill's route, but he had to. He

had to leave with his folks in such a hurry that he didn't have time to get anyone else to take the route; he just notified the paper manager. Well, when Bill got back from his vacation he had lost his paper route because Jimmy, the boy he'd gotten as a substitute, had to leave it, so they'd hired another boy.

Here is another story like the first one:

There was a boy named Johnny and he had a paper route too. His parents were going on vacation and he had to find someone to take the route for him so he could go too. He got one of his friends, a boy named Paul, to take the route for a week. Paul got paid $5.00 in advance for the week so he didn't pay much attention to how he did the job. He forgot some places some of the days and didn't deliver at all to them. Some other customers got angry because he didn't roll the papers tight and parts of the paper blew away or sometimes got wet. When Johnny got back he found that he had lost one customer out of every five.

Now which one of the friends that took the paper route was worse, the first one or the other one?
Why?

8. This girl was sitting around the house on a day off from school. Her Dad asked her to go down to the grocery store to get some things because he and the mother didn't have a chance to go themselves. The girl said that she'd go in a few minutes and then she said she'd go after while. Anyway, she never got around to going. No one else had been able to go, so they had a pretty poor supper that night. The father was angry at her. He figured out a few different punishments for the girl. The first one he thought of was of not letting her go out at all that week. The second one he thought of was to whip her. The third punishment that he thought of was this: The girl had a bicycle that they'd bought her and often asked her Dad, who was an A-1 mechanic, to do some work on it. So the father figured that the next time the girl asked him, he'd do just like the girl; he'd never get around to helping her.

Which one was the fairest?
Which of the three punishments was the worst?
Which one was the one that you think the father gave her?

9. Two guys about 10 years old, who lived in the same apartment house, got hold of some firecrackers last Fourth of July. Firecrackers are illegal, of course, but besides that, they both had fathers who worked nights. The firecrackers woke their fathers up and each father took his own boy into the house. The first father gave his boy a couple of pretty hard swats for waking him up. The second father took his son in and just told him how he'd been up all night and needed to sleep, and that waking him up would be just the same sort of thing as if someone kept the boy awake at night after he wanted to go to sleep. So both boys went outside again and played more quietly. But there were still firecrackers left and the next day one of the boys set them off, waking up the two fathers again.

Which boy do you think it was this time?
Why?

10. There was a school dance and one small group was smoking in the school building, which was against the rules. The grownups found out that this had happened, but they couldn't find out who had done it, because no one would tell. Since they couldn't find out who the individuals were, the principal decided to discontinue the dances for the whole school for the rest of the semester.

Was this right?
Why?

11. Two fellows were out one night and wanted to go riding in a car. They didn't have one, but they saw a car that was parked in front of a house with the keys in it, so they took that car. They drove it around for awhile but then they spotted a police car so they stopped the car and got out and started to run through back yards and alleys to ditch the police. They ran for awhile but then one of them got winded and police caught up with him. The other one ran for awhile more and then he crossed a ravine by an old bridge. The bridge was rotten and a plank broke on him so that one leg went through and got broken. Now what do you think?

If this boy had not stolen the car and had been running across the ravine on the old, rotten bridge all the same, would he have still broken through and hurt himself?
Why?

12. This story is about a guy who came from a tough neighborhood. He did pretty well, generally, but he sometimes took things. He had a friend who was a very good basketball player but whose folks didn't have much money so that his buddy couldn't afford to get gym shoes. One day when this boy was downtown he went into a store and took some gym shoes that they had there and gave them to his friend. Then a couple

of weeks later he was hanging around the drug store and managed to take a big box of chocolates that only cost about one quarter as much as the gym shoes. He didn't tell anyone about them; just kept them hidden and ate them. Now he wasn't caught either time and no one ever found out where his friend got the gym shoes.

Now which time was the stealing worse?

Why?

13. A boy was out with a group of friends one evening. They got to wrestling and one boy began wrestling each of the others, even when they didn't want to. He got to wrestling with a boy who was wearing glasses and the glasses fell off and broke. Now there were all sorts of things that the grownups could do to him; they could give him a spanking or they could have him work to pay for replacing the glasses or they could break something of his.

What would be the worst punishment?

What would be the fairest punishment?

Which one do you think that they gave him?

14. A girl wanted to use her mother's fur coat but the mother didn't want her to and told her not to use it because she was afraid the girl would maybe lose it in some way. The girl used it anyway figuring she wouldn't get caught, but the mother caught her. When she caught her using the coat she was so angry she made the girl stay in at night for a week.

There was another girl who wanted to use her mother's coat too, even though her mother didn't want her to and told her not to use it. The girl did use it anyway and the mother caught her at it but what this mother did was talk to the girl, explaining why she didn't want her to use it so that the girl would understand.

Now one girl used her mother's fur coat the next time she thought the mother was not around and the other one didn't use it again.

From these stories, tell me which one you think used the coat and which one didn't.

Why?

15. A science class was on a trip in the spring and they stopped to look over an apple orchard that was in bloom. While the leader wasn't around, one person in the group cut off many limbs, to get apple blossoms. He did this when none of the others were looking, so that none of the group knew who did it. The farmer saw the cut branches in the cars and was very angry. He complained to the group leader but no one knew who did it so no one could tell the group leader who had done it. The farmer and the group leader decided to charge everyone in the group equally to pay the damages, since they couldn't find out who had done it.

Was this right?

Why?

16. A girl named Mary didn't like another girl because this girl was going with a boy Mary liked. Mary called up this girl's mother and told her a lot of lies about the boy, so that the mother wouldn't let her daughter go out with him. Because of this Mary got to go out with him. The first time that they went out, another car hit the one they were in, and Mary got a broken nose.

Now if she'd have not told any lies, but had gotten to go out that same night with this boy she liked anyway, would she still have gotten a broken nose?

Why?

17. There was a young fellow named George who was working in a filling station. He was putting gas in a car and the driver asked him to check the oil. Then the driver went inside. George got done putting in gas and was just about to check the oil when another car came in that wanted service in a hurry. He went to wait on this car and then had to help with another car. He forgot all about checking the oil in the car he had started out with. The driver paid him but didn't ask about the oil because he figured George would have added some if it was needed. It happened that the oil was very low when George forgot to check it and soon there wasn't any oil left, so that the motor was damaged fairly much.

Here is another story about like the first one:

This young man's name was Dale and he worked as a station attendant, too. He was putting gas in a car when the driver asked him to check the oil. The driver went inside and then Dale got to talking to some of his friends who'd come in to see him. He figured that probably the car didn't need any oil anyway. When the driver came out he asked Dale how the oil was and Dale said that it was all right, even though he knew he hadn't checked it. The oil was low on this car, too, but the owner had it checked again before it ran completely out of oil, so that not too much happened to hurt this car.

Now which station attendant was worse?

Why?

18. This group of boys went into a candy store every day after school. When they started out,

everything was all right, but soon they began to take candy and things. One day the owner caught them. They admitted that they'd been taking things for quite awhile. The owner didn't know whether to call the juvenile squad of the police or to give the boys a spanking or to have the boys do things like sweeping and mopping the store till they'd worked to pay for what they had taken.

What punishment would be the fairest?
What punishment would be the hardest?
What did the storekeeper do?

19. Two boys got caught stealing a car. Their homes, their school records, etc., were exactly alike, and neither had led the other into taking the car, but for some reason one was sent to Redwing Training School and the other was put on probation and was allowed to stay at home. The one who stayed at home had a good probation officer who worked with him a lot. The one who was sent to Redwing was given lots of discipline till he was released after about a year and came home. Now they were both in their neighborhood again. About six months later one of them stole another car.

Which one do you think it was?
Why?

20. Some girls borrowed a car one night from the older brother of one of them. They went for a ride in the country and, since they all had drivers' licenses, they took turns driving. They all drove carefully but one was driving a little fast and skidded on the ice. The car went into the ditch and got a banged up grille and right fender.

Now was the one driving responsible for the accident or should they all be held responsible?
Why?

In questions 1, 6, 11, and 16, all dealing with immanent justice, the mature response would be one which brought in the idea of chance—that chance caused one boy to escape while the others were caught. The immature response would be the one that attributed some sort of supernatural cause to these events.

Questions 2, 7, 12, and 17 all had to do with moral realism. The mature response would be one in which acts were judged in terms of intent; the immature response would judge an act according to its consequences.

The problem of choosing between punishment in terms of expiation or retribution as opposed to punishment in terms of restitution appears in questions 3, 8, 13, and 18. The mature response would be the restitutive one, chosen because it was fairest, not because it was hardest. (Sometimes the restitutive punishment might be chosen, not because it was fairest, but solely because it was considered to be the most severe. In this case the response would be considered an immature one.) The choice of an expiatory punishment would be scored as an immature response.

Questions 4, 9, 14, and 19 had to do with the efficacy of severe punishment. Answers in which the position was taken that the less severe punishment was more effective were scored as mature, while those responses that indicated a belief in the efficacy of the more severe punishment were scored as immature.

Communicable responsibility was dealt with in questions 5, 10, 15, and 20. In each case, those responses that rejected communicable responsibility were judged to be mature, while those that accepted the idea of communicable responsibility were considered to be immature. This is not in strict accordance with Piaget's position. He says that both very young and older children accepted communicable responsibility while the middle age group of his sample rejected it. Very young children accept communicable responsibility because they believe that everything an adult does is fair while older children accept communicable responsibility as evidence of peer group solidarity. Since there is cross-cultural evidence (e.g., that gathered by Durkheim and by Fauconnet) that advanced cultures tend to reject communicable responsibility, the writer decided to score the rejection of communicable responsibility as the mature response.

Ninety-three student judges first categorized the items, separating them into appropriate areas. Their percentage of success in proper categorization was 82 per cent. Since few, if any, of these students had heard of Piaget's work, except in two periods of lecture, the success of these placements would suggest that the questions are readily associated with the specific areas of moral judgment that they are designed to measure. If some questions were so difficult to comprehend that they were answered randomly, one could expect that extraneous factors, not directly related to moral judgment, would produce most of the variance. Therefore, the student judges rated the areas of moral judgment with reference to which area they would expect a change from immature to mature moral judgment to occur first. No area was chosen as the one in which change would occur by more than 29 per cent of the sample.

The judges were also asked to select the easiest and the most difficult individual questions. No single question was chosen as the easiest by more than 10 per cent of the judges, and only three questions were considered most difficult by over 10 per cent of the judges. These were question 1, chosen as most difficult by 12 per cent of the judges, question 2, chosen as most difficult by 14 per cent, and question 20, chosen as most difficult by 17 per cent. Thus, no question seemed exceptionally easy or difficult to any large proportion of the judges, and no area of questions seemed, to these judges, to be of such general simplicity that a change toward mature belief might occur sooner than in other areas.

Successful categorization by the judges indicates that these items were directed toward the five aspects or areas of moral judgment to be studied, as these areas are defined by Piaget. The items were of about equal difficulty, in the opinion of adult judges. The actual reliability of this scale could not, of course, be determined until it was administered to a large sample of children. The extent of relation found for all other measures, whether of the relation between antecedent conditions and moral judgment, of the relation between moral judgment and behavior, or between the items themselves, is limited by the reliability of this test instrument.

A test of the relation of various antecedent conditions to the system of moral judgment used by the subjects involved the use of other measuring devices, plus the gathering of information from school records. The antecedent conditions were adult constraint, egocentricity, age (as measured by grade in school), sex, parental occupation, intelligence quotient, and religious training. Information concerning all but the first two of these variables was obtained from school records.

Many of the investigators working in the area since the time of Piaget's publication have believed adult constraint to be most significant in producing variance in the moral judgment of children. This belief seems to be based on the finding, common to many studies, that there is considerable social class difference in the age when a shift toward a more mature moral judgment occurs, with upper and middle socioeconomic groups shifting early and lower occupational groups changing in type of belief considerably later. It is assumed in the studies cited above that this difference is a result of the fact that lower occupational groups are more coercive and constraining. This interpretation seems open to question. Other factors, such as class differences in intelligence, might equally well account for the differences obtained. This study represents the first attempt to relate parental constraint to moral judgment by means of individual tests administered to a selected sample of the parents of the subjects.

The test of parent attitudes used was that devised by Shoben (1949). Shoben attempted to obtain information concerning the attitudes of the parents of problem and of nonproblem children. Seventy-four of his items discriminated between the two groups of mothers and could be placed in one of three clusters. (A few other items, not used in this study, discriminated, yet could not be placed in any specific grouping.) The clusters were relatively independent of one another. They were labeled "Ignoring," "Possessive," and "Dominating" and could have been scored (although Shoben did not do so) as independent subscales. Each item consisted of a statement (e.g., "A child should be seen and not heard") plus the choices of "strongly agree," "mildly agree," "mildly disagree," and "strongly disagree," following the statement. Scoring in Shoben's sample was such as to discriminate between the responses of mothers of problem and of nonproblem children. The scores ranged from those most characteristic of mothers of problem children down to those most characteristic of mothers of nonproblem children. Shoben had weighted the possible responses to each question according to differences between the proportions of mothers of problem children and the proportions of mothers of nonproblem children choosing them. The scoring system devised for this study was designed to measure the sheer amount of dominativeness, ignoringness, and possessiveness expressed in responses to questions in the subscales concerned with these three facets of parental attitude. The investigator and two additional judges (both psychologists) independently weighted the responses to each question in each of the three scales. For example, it was assumed that the person who strongly agrees that "children should give their parents unquestioning obedience" is more dominative than the one who strongly disagrees with this statement and that parents who believe that "children should be seen and not heard" are more ignoring than those who do not believe this to be so. The first question given above would be scored for dominance as follows: strongly agree, 4; mildly agree, 3; mildly disagree, 2; and strongly disagree, 1.

All three of the judges agreed on the weighting of responses to all but six of the 74 items. The weights assigned to responses on which there was a difference of opinion were those that were assigned by two of the three judges.

The other variable of significance to Piaget is that of developmental changes in thought processes. According to Piaget, the stage of egocentricity begins in the child in the time interval following the formation of sensorimotor intelligence. The stage of egocentricity includes, from the ages of 7 or 8 to 11 or 12, a period of concreteness, ". . . i.e., operational groupings of thought concerning objects that can be manipulated or known through the senses. Finally, from 11 to 12 years and during adolescence, formal thought is projected and its groupings characterize the completion of reflective thought" (1950, p. 123).

The egocentricity of the young child presumably predisposes him toward an immature system of moral judgment. The problem of measuring egocentricity seems a most difficult one. This writer is not familiar with any direct tests measuring all aspects of "egocentricity." . . . There is, however, a test available in one area that bears on Piaget's concept of egocentricity. This test is the Gorham Proverbs Test (1956) designed to measure concreteness and abstractness of thought. The stage of concreteness, for Piaget (1950), is a late portion of the egocentric period of the child's development. This stage is followed by a period characterized by increasing ability in "reflective thought." This "reflective thought" seems analogous to what is usually called the ability to deal with abstract materials. The Gorham Proverbs Test was designed to measure differential quantities of abstract as opposed to concrete thought. Therefore, the test appears to be measuring a major, late developing aspect of egocentricity.

The Shoben scale and the Gorham test are of somewhat limited utility in the study of factors involved in producing change in moral judgment. The Shoben test is a measure of attitudes, not of behavior. The greatest difficulty in the use of the Gorham test is that it does not measure egocentricity as would be desired, but measures concreteness (only one aspect of egocentricity) and abstractness (a characteristic that is dependent upon a lack of egocentricity). The writer could find no better tests suitable for group administration. The defects associated with these measuring devices force one to be cautious in drawing conclusions from the results obtained from these tests.

Procedure

The test of moral judgment was administered to the subjects in May of 1957. When presenting the moral judgment questions to the subjects in their classrooms, the writer introduced himself, passed out the test booklets, and then made the following statement:

> This is a sort of test, but it is not a test in the same way that you might be tested in arithmetic, because, as far as I know, there are no right or wrong answers. No one from the school here will know what you wrote down. The reason that I am asking these questions is this: I am interested in what young people think, because if we know more about what you think, we might be able to understand each other better. There are no right or wrong answers to these questions, so I d just like you to put down what you think or really feel. I'll read each question while you follow along on your copy of the questions. Then you put down your answer. When you are ready I'll read the next question and we'll go on that way. Don't look at your neighbor's paper; just put down what you think. Be sure and answer every question. Are there any questions? (If so, they were answered.) O.K., let's begin.

The questions were then read aloud to the subjects, who followed on their own test booklets, reading to themselves. Administration was not difficult. Students gave close attention and seemed to consider the questions significant and interesting. Only five sets of responses had to be discarded because the subjects clearly did not understand the directions or could not respond in terms of the directions—one indication of student attention and cooperation. Only one set of responses had to be discarded because the questions were obviously answered facetiously. Administration took 40 to 50 minutes.

Two class sections took the Gorham Proverbs Test at each grade level in grades 7, 9, and 11. Fifty-one students were tested in grade 7, 61 students in grade 9, and 55 students in grade 11. The testing schedule omitted grade 5 because the school officials did not wish to take time in this grade nor did the writer strongly urge the testing of this grade, for the Proverbs test, despite the fact that it had been standardized on populations as young as fifth graders, seemed quite difficult for the seventh grade sample.

The moral judgment test and the Proverbs test were administered in the school. Further data gathering involved contacting the parents of the subjects. The writer wished to determine the relations of adult constraint to abstractness-concreteness and to moral judgment. This required obtaining some measure of adult attitudes toward freedom and constraint. After repeated telephoning plus a number of home visits, the writer contacted the parents of 138 of the original sample of 167 subjects. In two cases, the people at the home could not talk English so these were dropped, leaving a total of 136 cases. Of the 136, five parents refused to take part, leaving a total of 131 cases where parents agreed to cooperate. Three pairs of siblings were in the sample of subjects tested on both abstractness-concreteness and moral judgment, so that only 128 sets of parents were involved in the parent attitude study. Of the 128 parents, 96 sent back their filled-out questionnaires. A comparison of the occupational levels of the respondents and nonrespondents (including those that could not be contacted) showed that 59 per cent of the parents in groups I to III responded, as compared with 61 per cent of the parents in groups V to VII.

It seems clear that the returns were not biased by disproportionate returns at different levels of parental occupation. Since this would be a principal area where one would expect to find bias, it seems probable that the responding parents were reasonably representative of the subject population.

Scoring

The Proverbs test was scored according to Gorham's protocols. The parent attitude survey was scored on the basis of parental ignoringness, possessiveness, and dominance. It should be noted that one additional measure of parent attitudes resulted from the scoring of the parent attitude questions. Extremeness of point of view —that is, how often the parent either strongly agreed or strongly disagreed rather than mildly agreed or disagreed—seemed related to the maturity of the responses made by that parent's child. This writer, therefore, scored each test for the total number of extreme positions taken in the entire test.

All answers to moral judgment questions were of an either-or or of a multiple choice type. Thus, it was a simple matter to score the moral judgment questions to indicate relative maturity of response. . . .

In general, the administration and the scoring of the tests was not difficult. The presentation of alternatives made the scoring of moral judgment questions simple and reliable and did away with much of the ambiguity and difficulty usually found in tests of this sort. Requiring a defense of the alternative response chosen probably led to somewhat greater interest and concern on the part of the subjects and permitted a check on the frequency of random and of facetious responses. The major test instrument, the moral judgment test, was notably successful, both in arousing interest and cooperation and in ease of scoring.

RESULTS

The data to be presented in the following pages are divided into sections. There are a number of problems to be dealt with, and, for the purpose of clarity, they will be taken up in serial order. The first four sections of the results and discussion deal with responses to the moral judgment questions themselves, and the fifth with the relation of moral judgment to antecedent conditions. The results appear in the following order: (a) reliability of the moral judgment test; (b) intercorrelations of responses to moral judgment questions; (c) correlations of responses within moral judgment areas; (d) correlations of responses between moral judgment areas; and (e) relation of moral judgment responses to antecedent conditions.

Reliability of the Moral Judgment Test

Information was obtained concerning the reliability of the total test (20 items) and also of the four-item subtests in each of the five moral judgment areas (immanent justice, moral realism, retribution vs. restitution, efficacy of severe punishment, and communicable responsibility) at the four age levels tested. The results are presented in Table 1.

It is apparent that the reliability of the 20-item moral judgment test, and also of the subscales, was not as high as that usually attained in educational tests. Subscale reliabilities were sometimes higher than the reliability of the whole scale, indicating that the various areas of moral judgment are not as closely knit as one would expect on the basis of Piaget's discussion. . . . Statistically, the relatively low reliability values of the subscales and of the whole scale set definite limits to the size of the inter-

Table 1: *Reliability of moral judgment questions*

	Grade 5	Grade 7	Grade 9	Grade 11
Total Test	.61	.59	.55	.56
Immanent Justice	.66	.68	.67	.67
Moral Realism	.58	.49	.30	.38
Retribution vs. Restitution	.45	.17	.28	.15
Efficacy of Severe Punishment	.60	.59	.50	.44
Communicable Responsibility	.32	.46	.39	.26

Table 2: *Correlations between responses having to do with immanent justice*

	Question 6	Question 11	Question 16
Question 1			
Grade 5	+.17*	+.38**	−.01
Grade 7	+.43**	+.17*	+.07
Grade 9	+.59**	+.40**	+.55**
Grade 11	+.27**	+.40**	+.42**
Question 6			
Grade 5		+.79**	+.74**
Grade 7		+.63**	+.74**
Grade 9		+.73**	+.81**
Grade 11		+.81**	+.77**
Question 11			
Grade 5			+.78**
Grade 7			+.71**
Grade 9			+.74**
Grade 11			+.78**

* *Significant at the .05 level.*
** *Significant at the .01 level.*

relations that can appear among the so-called varieties or areas of moral judgment.

Intercorrelations of Responses to Moral Judgment Questions

Responses to each item on the 20-item test were correlated with responses to every other item at each of the four age levels—a total of 760 correlations. Among this many correlations, where no relation of any sort actually existed, one would expect to find, by chance, 38 values significant at the .05 level and beyond. There actually were 294 positive correlations and 79 negative correlations significant at or beyond the .05 level of confidence. It seems clear that responses to moral judgment questions were positively correlated with one another to a far greater extent than one could expect by chance. A larger number of negative correlations were obtained than one could expect by chance. Some questions that seem, logically, to be positively related to other questions were apparently negatively correlated with other questions at one or more grade levels if judged in terms of the responses obtained.

Correlations of Responses Within Moral Judgment Areas

Questions 1, 6, 11, and 16 of the moral judgment test measured immanent justice; 2, 7, 12, and 17, moral realism; 3, 8, 13, and 18, retribution vs. restitutive punishment; 4, 9, 14, and 19, the efficacy of severe punishment; and 5, 10, 15, and 20, communicable responsibility. Correlations between responses in the various moral judgment areas are presented in Tables 2, 3, 4, 5, and 6.

Table 3: *Correlations between responses having to do with moral realism*

	Question 7	*Question 12*	*Question 17*
Question 2			
Grade 5	+.42**	+.33**	+.35**
Grade 7	+.33**	+.18*	+.24**
Grade 9	+.45**	+.12	+.23**
Grade 11	+.64**	+.19**	+.57**
Question 7			
Grade 5		+.41**	+.53**
Grade 7		+.23**	+.47**
Grade 9		+.22**	+.38**
Grade 11		+.21**	+.57**
Question 12			
Grade 5			+.47**
Grade 7			+.35**
Grade 9			+.53**
Grade 11			+.11

* *Significant at the .05 level.*
** *Significant at the .01 level.*

Table 4: *Correlations between responses having to do with retribution vs. restitution*

	Question 8	*Question 13*	*Question 18*
Question 3			
Grade 5	+.21**	+.35**	+.21**
Grade 7	+.19**	+.37**	+.07
Grade 9	+.41**	+.58**	+.24**
Grade 11	+.16*	+.60**	+.37**
Question 8			
Grade 5		+.36**	+.39**
Grade 7		+.29**	+.17*
Grade 9		+.15*	+.14*
Grade 11		+.17*	+.25**
Question 13			
Grade 5			+.26**
Grade 7			+.41**
Grade 9			+.35**
Grade 11			+.24**

* *Significant at the .05 level.*
** *Significant at the .01 level.*

Of the correlations of responses within areas of moral judgment presented in Tables 2 to 6, 102 out of 120 correlations were positive and significant. The area of immanent justice showed higher correlations between responses than other areas, but correlations in all areas of moral judgment were such that one is justified in talking about consistency within areas of moral judgment, especially when considering the reliability of the test items. Items 1 and 4 seemed less closely related to their respective areas of moral judgment than were most items. Item 20, an attempt to push the rejection of communicable responsibility to its logical extreme, was more often negatively than positively correlated to other items dealing with this aspect of moral judgment in grades 5 and 11; on the other hand, it was more often positively correlated with other items in this area in grades 7 and 9. It seems that one can legitimately discuss areas of moral judgment, that these areas exist, since responses to the different questions within areas were almost always correlated positively and significantly with one another.

Table 5: *Correlations between responses having to do with efficacy of severe punishment*

	Question 9	*Question 14*	*Question 19*
Question 4			
Grade 5	+.09	+.16*	+.30**
Grade 7	+.09	+.12	+.13
Grade 9	+.22**	+.19**	+.45**
Grade 11	+.23**	+.24**	+.16*
Question 9			
Grade 5		+.73**	+.58**
Grade 7		+.81**	+.61**
Grade 9		+.57**	+.51**
Grade 11		+.76**	+.53**
Question 14			
Grade 5			+.57**
Grade 7			+.44**
Grade 9			+.46**
Grade 11			+.43**

* *Significant at the .05 level.*
** *Significant at the .01 level.*

Table 6: *Correlations between responses having to do with communicable responsibility*

	Question 10	*Question 15*	*Question 20*
Question 5			
Grade 5	+.16*	+.51**	−.20**
Grade 7	+.54**	+.29**	+.05
Grade 9	+.36**	+.25**	+.20**
Grade 11	+.19**	+.28**	−.20**
Question 10			
Grade 5		+.38**	−.51**
Grade 7		+.48**	+.18*
Grade 9		+.30**	+.03
Grade 11		+.28**	−.14*
Question 15			
Grade 5			−.26**
Grade 7			+.02
Grade 9			+.01
Grade 11			+.01

* *Significant at the .05 level.*
** *Significant at the .01 level.*

Correlations of Responses Between Moral Judgment Areas

A more basic problem is the determination of whether there is any consistent correlation between the type of response made in one area of moral judgment and responses made in each of the other areas of moral judgment. It is implicit in Piaget's theory (1950, pp. 106–107) that response tendencies in the various areas of moral judgment (excluding communicable responsibility) are positively related to one another. This is a testable proposition. . . . The correlations of response tendencies between moral judgment areas are presented in Table 7.

Twenty of the 40 correlations between areas of moral judgment were significant. . . . All of the significant correlations shown in Table 7 were positive, except for the correlation of moral realism with communicable responsibility at grade 7, where the relation was negative, and the correlation of immanent justice with efficacy of severe punishment at grade 11, where the relation was curvilinear. In this curvilinear relation, both highly mature and highly immature sets of responses to the immanent justice question were positively related to mature responses to questions concerning the efficacy of severe punishment.

The intercorrelations among areas of moral judgment seem to show that mature moral judgments in the areas of moral realism, retribution vs. restitution, and efficacy of severe punishment were the most closely correlated with one another, that responses to questions concerning immanent justice were somewhat less closely correlated to responses in other areas of moral judgment, and that belief or nonbelief in communicable responsibility was even less closely related to other aspects of moral judgment.

Relation of Moral Judgment Responses to Antecedent Conditions

Piaget suggests that changes from immature to mature moral judgment result from changes in the amount of adult constraint and peer group cooperation experienced by the child and from qualitative changes in the child's method of thought. The child's level of development, for Piaget, is best predicted from chronological age. Most students of moral judgment since Piaget have found intelligence and social class to be related to the type of moral judgment made by the child. All of these variables, except peer cooperation, were measured in this study. Each significant relation between these variables and moral judgment responses is presented in Table 8.

From the data in Table 8 it seems that IQ and parental occupation were more closely related to moral judgment than were the other variables studied in the areas of moral realism, retribution vs. restitution, and the efficacy of severe punishment. In general, one can say that brightness, represented by IQ, was more closely related to the type of moral judgment used in those three areas listed above than was any of the other variables measured in this study. Since parent's occupation and child's IQ were themselves positively correlated, it may well be that the correlations found between parent's occupation and the moral judgment of the child were merely reflections of differences due, basically, to IQ, rather than to cultural differences denoted by occupational level. The results of a determi-

Table 7: *Correlations of responses between moral judgment areas*

	Moral Realism	*Retribution vs. Restitution*	*Efficacy of Severe Punishment*	*Communicable Responsibility*
Immanent Justice				
Grade 5	.12	.27**	.09	.13
Grade 7	.29**	.18*	.17	.06
Grade 9	.16*	.09	.14	.12
Grade 11	.07	.19*	.27**	.09
Moral Realism				
Grade 5		.23*	.34**	.30**
Grade 7		.33**	.32**	.18*
Grade 9		.20*	.33**	.15
Grade 11		.16	.10	.00
Retribution vs. Restitution				
Grade 5			.18*	.16
Grade 7			.33**	.18*
Grade 9			.30**	.00
Grade 11			.22*	.09
Efficacy of Severe Punishment				
Grade 5				.21*
Grade 7				.06
Grade 9				.08
Grade 11				.04

* *Significant at the .05 level.*
** *Significant at the .01 level.*

nation of class differences in parental attitudes, discussed later in this study, seem to favor this position.

Immanent justice and communicable responsibility responses were less often significantly related to the antecedent conditions of sex, parental occupation and IQ than were judgments concerning moral realism, retribution vs. restitution, and the efficacy of severe punishment. It may be that the higher degree of association between responses to moral judgment questions in the latter three areas is a result of the fact that response tendencies in these areas were related to the same antecedent conditions.

Chronological age was positively and significantly related in all areas of moral judgment —even "communicable responsibility" where, from Piaget's analysis, one would expect a negative relation in this age range. Age change was by no means saltatory, however—no sudden shifts in response tendencies were evident.

Out of 45 correlations of abstractness with moral judgment, six were significant. If abstractness is indicative of a freedom from egocentricity, one would expect high abstractness scores to be positively related to mature moral judgment. Five of the six significant correlations were in this direction. Only three of the 45 correlations of concreteness with moral judgment were significant. If concreteness is a manifestation of egocentricity, concreteness should be negatively correlated with mature moral judgment. The reverse held true in two of the three significant correlations.

If one scored the parent attitude scale in terms of the amount of ignoringness, possessiveness, and dominance expressed by the parents, one would expect, from Piaget's theory, to obtain the following results: belief in immanent justice, moral realism, retribution, and expiation, as opposed to restitution or reciprocity, and in the efficacy of severe punishment are all learned from the parents. If this is so, possessiveness and dominativeness should be positively correlated and ignoringness negatively correlated with acceptance of these beliefs. On the other hand, acceptance of communicable responsibility is usually learned from the peer group, which uses

Table 8: *Significant correlations of antecedent factors with moral judgment responses*

Factor	*Immanent Justice*	*Moral Realism*	*Retribution vs. Restitution*	*Efficacy of Severe Punishment*	*Communicable Responsibility*
IQ	.17(9)* .23(11)†	.34(5)* .36(7)* .31(9)* .20(11)*	.31(5)* .42(7)* .23(9)*	.24(5)* .30(7)* .30(9)*	
Parent Occupation	.21(9)‡ .22(11)‡	.21(7)‡ .17(9)‡ .16(11)‡	.23(5)‡ .23(7)‡	.18(5)‡	
Age (entire sample combined)	.35§	.35§	.25§	.26§	.12§
Abstractness	.29(9)¶ .26(11)¶	.25(7)¶	.26(7)¶	.34(9)¶ .46(11)‖	
Concreteness		.31(11)**		.36(7)††	.38(9)††
Parent Attitudes					
Ignoringness	.31(9)§§		.39(7)‡‡		.32(7)‡‡ .32(7)‡‡
Possessiveness	.53(11)§§				.52(11)‡‡ .40(11)‡‡
Dominativeness				.44(11)§§	
Extremeness of Point of View	.32(11)§§			.40(9)§§	.29(7)§§ .40(11)§§

Note.—Grade level at which correlation was found is indicated in parentheses following each coefficient.

* *High IQ positively related to mature judgment.*

† *IQ related to moral judgment curvilinearly; highest and lowest IQ made most mature responses.*

‡ *Higher parental occupation positively related to mature moral judgments.*

§ *Increasing age positively related to mature moral judgments.*

¶ *High abstractness positively related to mature moral judgments.*

‖ *High abstractness negatively related to mature moral judgments.*

** *High concreteness negatively related to mature moral judgments.*

†† *High concreteness positively related to mature moral judgments.*

‡‡ *High amounts of the parent attitude positively related to mature moral judgments.*

§§ *High amounts of the parent attitude negatively related to mature moral judgments.*

communicable responsibility as an evidence of group solidarity. Here it would seem probable that the ignored child, more free to interact with peers, should show less mature moral judgment than the child of possessive or dominative parents who presumably would have somewhat fewer peer group contacts. Of the 45 correlations of the parent attitudes of expressed dominativeness, possessiveness, and ignoringness with the level of the child's moral judgment, eight were significant, seven of them in the direction predicted from the above discussion.

Extremeness of point of view on the part of the parent was correlated more frequently with the moral judgment of the child than were Shoben's subscales of ignoringness, possessiveness, or dominativeness. Extremeness of point of view was not significantly correlated with the type of parental response to any of the three subscales; there was no significant correlation, for instance, between the amount of dominativeness expressed and the number of extreme positions taken—so that the parent's extremeness of point of view seems, in its own right, to be a significant factor in relation to the child's level of moral judgment. The parents who more frequently take extreme points of view tended to have children who are immature in their moral judgments.

The existing significant correlations of parent attitude to moral judgment were not distributed randomly, but occurred most frequently in the areas of immanent justice and communicable responsibility, the very areas that showed few significant correlations with such variables as intelligence and parental occupation.

One can ask whether parent attitudes as measured here may be merely a reflection or indirect measure of social class differences in approach, either to childrearing or to answering questionnaires. The writer attempted to determine whether this might be the case. The responding parents had already been categorized into occupational groups. They were now divided into two larger groupings, those in classes I to III and those in classes V to VII of the Minnesota Occupational Scale. Their responses were divided into those above and below the median in amount of expressed dominativeness, possessiveness, and ignoringness and in the number of extreme points of view. A comparison of the attitudes of these two groups allows one to determine whether the obtained correlations between parent attitudes and moral judgment are actually a result of some occupational group bias in responses. An analysis of the data indicates that this was not the case. There were no significant differences in attitude between the two groups. Parent attitudes, as expressed on the Shoben scale, were not significantly related to parental occupation as measured by the Minnesota Occupational Scale.

DISCUSSION

The primary purpose of this study was to determine the amount of association between responses to moral judgment questions.

The reliability of the test items, as a whole scale, seems to indicate in itself that responses concerning various aspects of moral judgment are related to one another empirically as well as logically, but not to as large an extent as one would expect from Piaget's analysis. Reliability did not change appreciably from one age to another. This seems to indicate (as do the actual correlations between responses) that moral judgment remains consistent in the amount of association from one area to another, contrary to what one might expect.

The reliability of subscales measuring specific areas of moral judgment was often somewhat higher than that of the total scale, suggesting a low association between areas but a higher association within areas. The reliability of the subscales seems rather high, in general, since each subscale consisted of only four items. Although high, when looked at in this fashion, subscale reliability was still so low that one cannot expect inter- and intra-moral judgment area correlations to be as high as they might be with a more reliable test. It should be noted, however, that the whole scale reliability of this test is equal to that of some widely used measuring devices. . . .

The correlations of responses within moral judgment areas, like the reliability figures, seem to argue for quite considerable consistency within each area. This consistency may be due, in a large part, to the common correlations of items within an area to such things as intelligence and socioeconomic differences.

The results tend to support the idea that there is at least some consistency in response between areas. The correlations between responses in the five different areas of moral judgment indicate that judgments made regarding moral realism, retribution and expiation vs. restitution and reciprocity, and the efficacy of severe punishment were often significantly related to one another. Immanent justice was somewhat less frequently correlated positively and significantly with these three areas, though certainly there were far more positive and significant correlations than one would expect by chance. Communicable responsibility was less closely correlated with other areas of moral judgment; in fact, one of the four significant correlations of communicable responsibility with other areas of moral judgment was negative. The fact that the data were obtained separately at each of four age levels with quite similar results lends support through replication to the findings discussed above.

Piaget assumes, as a result of his logical analysis and of his findings, that response tendencies are similar in all areas of moral judgment (except for communicable responsibility), that there is what we would now call a general factor of moral judgment. The findings above were, in general, the results one would expect on the basis of Piaget's theoretical discussion, except that responses to questions involving communicable responsibility seemed largely unrelated, rather than negatively related, to responses in the other four moral judgment areas. With this exception the correlations were in the expected direction but were considerably lower than one would expect from Piaget's own work. The reliability data, as well as the interarea correlations of the responses, seem to show that what might be called a general factor of moral judgment was present, but with a rather low degree of saturation.

The second purpose of this study was to determine the relation of a number of variables to the type of moral judgment used by the subjects. A number of antecedent conditions were investigated. Some of these factors, such as sex, IQ, and parent occupation were touched on peripherally in Piaget's investigation and discussion of moral judgment. Other factors were discussed more comprehensively by Piaget. These were age (as a sign or indicator of the condition of the organism with reference to the next two factors, rather than a direct cause of change), egocentricity (which the writer attempted to measure, admittedly indirectly, through a test of abstractness-concreteness), and adult constraint. The latter two factors, interacting with one another and with peer group cooperation, were believed by Piaget to be directly causal in producing changes in beliefs concerning moral judgment. Here, too, the results obtained in this study must be judged in terms of the reliability of the measuring instruments.

Of those variables that were not discussed to any extent by Piaget, IQ was clearly the most significant in its relation to responses concerning moral realism, retribution and expiation vs. restitution and reciprocity, and the efficacy of severe punishment. Sex of subjects and parent occupation of subjects also showed a considerable number of significant correlations with responses in these three moral judgment areas. Immanent justice seemed less closely related, and communicable responsibility seemed essentially unrelated to this group of variables.

Age was correlated with the type of response made in all five areas. One would expect on the basis of Piaget's work that increasing age is positively correlated with more mature moral judgment in all areas except that of communicable responsibility, where an increase in peer group solidarity might well cause a greater frequency of immature responses as age increased. Contrary to Piaget's theory, even communicable responsibility showed a very low but significant positive correlation between age and mature moral judgment. Within the age groups studied there was certainly no evidence for any large, sudden change in beliefs concerning moral judgment.

Parent attitudes, as expressed on questionnaire responses, most often showed significant correlations with responses to questions concerning communicable responsibility and immanent justice. These are the very areas that were less closely correlated with the first group of antecedent conditions. Abstractness and concreteness were related to moral judgment in the direction predicted by Piaget, but perhaps because abstractness is itself related to IQ. None of these variables was closely and consistently related to moral judgment to the extent that it accounted for the major portion of the variance, perhaps because the test of moral judgment was itself less reliable than one would desire.

This writer's general conclusions regarding these antecedent conditions are that IQ and chronological age were the variables most closely correlated to the type of response made with reference to moral realism, retribution and expiation vs. restitution and reciprocity, and the efficacy of severe punishment. Age and to some extent, IQ and parent attitudes, seemed to be correlated with responses to immanent justice items. Only parent attitude items seemed correlated to any extent with type of response made to questions involving communicable responsibility.

This study provided a number of moral judgment questions of known reliability and also information about their relation to one another and to various antecedent conditions. The questions had an additional value in that they evoked a very high degree of interest among subjects within the age groups studied.

It seems to this writer that the next step in research is to develop a more reliable moral judgment test and then to determine the relation of the various aspects of moral judgment to behavior. All the studies of moral judgment done since Piaget's time have limited themselves to testing some aspect of his theory. Knowledge of an individual's beliefs concerning the various areas of moral judgment might have practical as well as theoretical importance. This question, whether an understanding of any individual's concepts of moral judgment has any practical application, is one that has not yet been asked, much less answered. This is an area in which further research may fill a definite gap in our knowledge.

SUMMARY

Piaget studied children's moral judgments. His research showed that an age change occurred in moral judgment. Piaget attributed these changes to increased peer group cooperation, decreased adult constraint, and qualitative intellectual changes in the child as the child grows older. The primary purpose of this study was to determine the consistency of responses to questions within various areas of moral judgment (immanent justice, moral realism, expiatory vs. restitutive punishment, efficacy of severe punishment, and communicable responsibility) and between areas of moral judgment. The secondary purpose of this study was to determine the relation of various antecedent conditions to children's moral judgments.

The sample population consisted of 807 subjects, all of the children in grades 5, 7, 9, and 11 in a midwestern city. Subjects approximated U. S. urban norms in parental occupational level and in IQ.

A moral judgment test was administered to all subjects. The test questions were modeled after Piaget's, but were more difficult and were oriented to an older age group. A test of abstractness-concreteness was administered to subsamples at grades 7, 9, and 11. Parents of subjects in these subsamples responded to a parent attitude test.

The reliability of the entire moral judgment scale was approximately .60 at each age level. Correlations between moral judgment responses revealed far more positive and significant correlations than might be expected by chance. Responses within moral judgment areas (e.g., immanent justice) were nearly always positively and significantly correlated. Correlations of the number of mature responses in the various areas of moral judgment showed response tendencies in the areas of moral realism, retribution vs. restitution, and the efficacy of severe punishment to be rather closely related. Responses to immanent justice questions were less closely related, while responses to questions involving communicable responsibility were essentially unrelated to other response tendencies.

Abstractness and concreteness were only slightly related to moral judgment responses. Parent attitudes were significantly related to moral judgments, expecially in the areas of immanent justice and communicable responsibility. IQ and, to a lesser extent, parental occupation were positively and significantly correlated with mature moral judgment in all areas.

Most findings having to do with the correlations of responses to moral judgment questions were in the direction suggested by Piaget. Obtained correlations were lower than would be expected from his analysis, perhaps because the test of moral judgment was not highly reliable. Those factors measured in this study that Piaget believed most influential in producing change in moral judgment seemed of relatively slight significance, except that parental constraint seemed closely related to responses made in the areas of immanent justice and communicable responsibility.

REFERENCES

DENNIS, W. Animistic thinking among college and university students. *Sci. Mon.*, 1953, *76*, 247–250.

GORHAM, D. R. A proverbs test for clinical and experimental use. *Psychol. Rep.*, 1956, Monogr. Suppl. No. 1.

MEDINNUS, G. R. An investigation of Piaget's concept of the development of moral judgment in six to twelve year old children from lower socioeconomic class. Unpublished doctoral dissertation, University of Minnesota, 1957.

Minnesota scale for paternal occupations. University of Minnesota Press, 1950.

PIAGET, J. *The moral judgment of the child.* New York: Harcourt, Brace, 1932.

PIAGET, J. *The psychology of intelligence.* New York: Harcourt, Brace, 1950.

SHOBEN, E. J., JR. The assessment of parental attitudes in relation to child adjustment. *Genet. Psychol. Monogr.,* 1949, *39,* 101–148.

60

The Motivating Effect of Learning by Directed Discovery

Bert Y. Kersh

The role of discovery as described by Bruner (8) is subjected to test in this research by Kersh. Because the experimental conditions play a key role in determining the results, the validity of a theoretical construct may be subject to doubt because of the ways in which it was tested. A careful definition of terms and conditions is necessary before the reader can generalize beyond a given study, and such phrases as "learning must be meaningful" require critical examination rather than common-sense acceptance and attempts at application.

Advocates of the process of learning by directed discovery claim a number of advantages, most of which are included in a recent article by Bruner (1961).* He has suggested that learning by discovery benefits the learner in four ways: it (*a*) increases the learner's ability to learn related material, (*b*) fosters an interest in the activity itself rather than in the rewards which may follow from the learning, (*c*) develops ability to approach problems in a way that will more likely lead to a solution, and (*d*) tends to make the material that is learned easier to retrieve or reconstruct.

Research evidence does not entirely support Bruner's arguments. One of the more recent reviewers, Ausubel (1961, p. 47), concludes "that most of the reasonably well-controlled studies report negative findings." However, as is true in other areas of research, the evidence is somewhat equivocal, partly because it is difficult to equate studies in terms of the amount and kind of direction that is provided. The experimental subjects rarely if ever are required to learn completely without help, and the kind of help provided commonly differs. Consequently, there are studies which appear to be somewhat contradictory, such as Craig's (1956), in which the "directed" group learned and retained significantly more principles than the "independent" group, and Kittell's (1957), in which the group which received an intermediate amount of guidance was superior in learning, retention, and transfer to groups receiving either more or less direction. It has been suggested that the "intermediate" amount of guidance provided by Kittell may have actually exceeded the amount Craig provided to his directed group (Ausubel, 1961, p. 52).

One of the few studies that forced learners to discover almost entirely without help provides data in support of the discovery process (Kersh, 1958). The contrasting directed treatment groups were superior in learning rate and immediate recall, but the "no help" group was superior in terms of retention and transfer after a period of approximately one month following the learning period. No evidence was produced to indicate that the no help group understood the rules better. Instead, an explanation was offered in terms of practice. On the basis of a subjective analysis of the subject's comments written on the retests and reported to the experimenter, it was concluded that the learners were motivated to continue the learning process or to continue practicing the task after the formal learning period.

The present experiment was designed to provide formal data concerning the motivating power in question.

HYPOTHESIS

Each subject had the task of learning the following two rules of addition:

1. Odd Numbers rule. The sum of any series of consecutive odd numbers beginning with 1 is equal to the square of the number of figures in

---

From *Journal of Educational Psychology,* 1962, 53, 65–71.

* This article appears in this volume—*Editor's note.*

the series. (For example, 1, 3, 5, 7, is such a series; there are four numbers, so 4 × 4 is 16, the sum.)

2. Constant Difference rule. The sum of any series of numbers in which the difference between the numbers is constant is equal to one-half the product of the number of figures and the sum of the first and last numbers. (For example, 2, 3, 4, 5, is such a series; 2 and 5 are 7; there are four figures, so 4 × 7 is 28; half of 28 is 14 which is the sum.)

The rules can be learned by simple memorization of the task procedure as above. Further, the learner can become cognizant of certain relations which these rules bear to geometrical and arithmetical concepts, in which case it is assumed that his learning will be more meaningful. The definition of meaning as well as the geometrical and arithmetical relationships referred to are identified in a previous publication (Kersh, 1958). In the hypothesis statement below, the term "relationships" refers specifically to those in the reference cited above and generally to comparable relationships in related tasks.

As will be explained below, the experimental treatments in the present study differed primarily with respect to the extent of the external direction provided the subjects in learning the relationships referred to above. The present experiment was designed to test the following hypothesis.

To the extent that the external direction provided to the learner is lessened during his attempts to discover the relationships which are considered essential to the understanding of a cognitive task: (*a*) the learner will tend to use the learned material more frequently after the learning period (i.e., to extend the practice period voluntarily) and, as a result, (*b*) he will remember it longer and transfer his learning more effectively.

It should be noted that the hypothesis is written in two parts and that the second is dependent upon the first.

PROCEDURE

A total of 90 high school geometry students was utilized, having been selected from a larger group on the basis of a pretest covering the arithmetical and geometrical concepts and procedures that were considered essential prerequisites to the tasks used in the experiment. The entire sample was then taught the two rules of addition given above by being simply told the rules and given practice in their application. They were taught by a programed booklet procedure to the same criterion, six successive applications of each of the two rules. Thereafter, the subjects were divided at random into three main groups of 30 each, and each group was treated differently.

One group, called the Directed Learning group, was taught the rules and their explanation entirely by a programed learning technique. Each subject learned from a booklet in which the learning process was broken down into smaller steps, and answers to questions or solutions to problems were revealed to the subject whether he responded correctly or not.

A second group, called the Guided Discovery group, was required to discover the explanation with guidance from the experimenter. The subjects in the Guided Discovery group were taught tutorially using a form of Socratic questioning which required each subject to perform specific algebraic manipulations and to make inferences without help. The guidance was a practical expedient, since it was necessary to control between groups the quality and quantity of the relationships used in explaining the rules.

The final group was called, appropriately, the Rote Learning group since the explanation for the rules was omitted. This treatment was incorporated in the research design primarily as the control for "meaningful" learning.

After the learning period of the experiment, a test of recall and transfer was given to subgroups of each treatment group after 3 days, 2 weeks, and 6 weeks. For this purpose each of the three main groups was divided into three subgroups of 10 each.

The test consisted of two problems and a short questionnaire. The problems were given first with instructions to show all work including scratchwork. The two test problems were as follows:

1. John's employer agrees to pay him $1.00 for his first day of work and increase his pay by $2.00 each day. How much will he receive for the first month's work if he works all 30 days?

2. A man is left a sum of money by an eccentric relative. The will states that he will receive $10.00 the first month and that each successive monthly payment will be increased by $5.00 (i.e., he will receive $10.00 the first month, $15.00 the second month, $20.00 the third month, etc.). His monthly payment at the end of four years is

Table 1: *Number of subjects (of 10 in each cell) who used and stated the rules correctly on the retest*

Treatment Groups	Used Rules		Stated Rules	
	Odd Numbers 1	Constant Difference 2	Odd Numbers[a] 3	Constant Difference[a] 4
Rote Learning:				
3 days	7	7	7	9
2 weeks	7	6	2	6
6 weeks	4	4	0	3
Guided Discovery:				
3 days	6	6	8	9
2 weeks	3	5	3	4
6 weeks	2	3	3	3
Directed Learning:				
3 days	4	3	3	4
2 weeks	4	3	1	3
6 weeks	0	3	1	1

[a] *Differences between treatment groups and between test periods significant by chi square at or beyond .05 level.*

$245.00. What is the total amount he has been paid by that time?

The questionnaire asked the subject to state each rule, using examples if necessary, and to report whether or not he made use of the rules after the formal learning period.

RESULTS

The number of subjects in each group who used the appropriate rule in an acceptable way on the test was employed as the index of transfer. Acceptable use of a rule for the first test problem meant the use of either rule to obtain the solution; for the second test problem, only the Constant Difference rule was acceptable. Computational accuracy was not required.

The number of subjects in each group who wrote an acceptable statement of each rule was used as a measure of pure retention. To be acceptable, each subject's statement had to be complete and accurate, but not necessarily in the same words as the original statements. Errors in spelling or grammar were overlooked.

Table 1 presents the number who used and stated the rules in the acceptable way on the test problems. A total of 90 subjects served as the basis for the data in Table 1, 10 subjects per cell. . . .

None of the interaction effects was significant, indicating that the rate of forgetting did not differ significantly across the teaching treatment groups. . . .

Otherwise, as pointed out by the footnote references in Table 1, the differences between treatment groups and between test periods were found to be significant for all columns except that headed "Constant Difference 2," for which the observed differences were found not to be reliable.

Perhaps the most striking finding in the present study is that the Rote Learning group was found to be consistently superior in every respect to the other treatment groups. Although this completely unanticipated finding has no direct bearing on the hypothesis in question, it does nevertheless bear clearly upon the related question of "meaningful vs. mechanical" learning. This finding will be discussed in a subsequent section.

Strictly speaking, the hypothesis which the present experiment was designed to test involves only the Guided Discovery and Directed Learning treatments. To support the major hypothesis, the data should have shown that the subjects comprising the Guided Discovery group used the rules after the learning period more frequently than the subjects in the Directed Learning group; and, if so, that the former remembered and transferred the rules more effectively than the latter.

With respect to the frequency of using the rules after the learning period, the results do support the hypothesis. Although the number of subjects in each group who reported that they did use the rules was very small, the difference between the frequency patterns of the two groups in question is statistically significant. Eleven subjects of the 30 in the Guided Discovery group reported that they had used the rules as compared with two subjects in the Directed Learning group. In the Rote Learning group, six subjects of 30 reported in the affirmative.

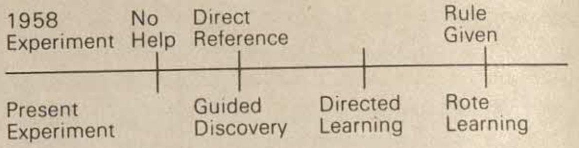
With respect to the relative permanence of the retention and increased transfer effects, the results also support the hypothesis. The Guided Discovery group is clearly superior to the Directed Learning group 3 days after the learning period, and since the rate of forgetting may be presumed to be approximately the same for each treatment group (see statistical analysis above), their initial superiority remains after 6 weeks.

DISCUSSION

The data from this present experiment do not support the generalization that learning by a process which involves discovery is necessarily superior to learning by more highly directed processes. Indeed, these data suggest that under certain conditions of learning, highly formalized "lecture-drill" techniques, ordinarily considered sterile and meaningless, produce better results than techniques which attempt to develop "understanding."

One explanation for the present results is that they reflect a simple and well known phenomenon, retroactive inhibition. The experimental efforts to inject meaning into the rules amounted to following their initial rote learning with a closely related and complex learning task; thus the Rote Learning group may have surpassed other groups simply because retention among the latter was inhibited by the interpolated learning.

How may the present results be reconciled with those of the previous experiment by Kersh (1958), in which learning by discovery proved markedly superior? The preferred interpretation is that the findings of the two studies are actually complementary. Schematically, the treatments employed in the two experiments may be compared on a line representing the continuum of learning processes: at one extreme, learning without any external direction whatsoever (true self-discovery); at the other, learning by lecture-drill processes (rote learning), as follows:



As is indicated above, the Direct Reference treatment in the 1958 experiment is comparable to the Guided Discovery treatment in the present one; similarly, the Rule Given and Rote Learning groups correspond. The present experiment has no counterpart to the No Help treatment of the previous study; and, in the previous one, the Directed Learning treatment was not represented.

When compared as above, the results of the two experiments are remarkably similar. The initial achievement of the comparable groups in both experiments was very high then dropped to where only about half of each group was able to recall and apply the rules after 4 to 6 weeks. In each experiment the difference in the performance of the Rote Learning and Directed Discovery groups was not notable; if anything, the Rote Learning groups tended to perform slightly better.

With respect to the motivating power of learning by discovery, in the 1958 experiment the superior performance of the No Help subjects on the retest together with their written comments and verbal reports to the experimenter strongly evidenced their increased interest. The present results leave no doubt that there is a tendency for interest to accrue as a result of learning by discovery.

The results of both experiments also are consistent in their failure to support the notion that attempts to provide added meaning will necessarily prolong memory for rules and procedures and will enhance their transfer. On the contrary, both experiments suggest that such attempts may well do more to interfere with learning than enhance it. This does not mean that rote learning is superior to learning with understanding. Rather it means that we need to know much more than we do about meaningful learning and how we come by it.

The relatively poor showing of the Directed Learning group in the present study is partially explained by the subjects' reported failure to practice the rules after the learning period to the extent that the subjects did in other groups. Why the Rote Learning treatment generated more interest than the treatment in question again may reflect nothing more than that the original learning was inhibited by the interpolated programed learning. The subjects' unfamiliarity with the instructional procedure may have contributed to their confusion.

Most certainly the data from the two experiments under discussion suggest that the frequently taught principles of learning that pertain to self-discovery and meaning (see introduction) should be restated or qualified. The following statements are offered for further study.

Learning by self-discovery. Learning by self-discovery is superior to learning with external direction only insofar as it increases student motivation to pursue the learning task. If sufficiently motivated, the student may then continue the learning process autonomously beyond the formal period of learning. As a result of his added experience, the learner may then raise his level of achievement, remember what he learned longer, and transfer it more effectively. The explanation for the elusive drive generated by independent discovery is not evident, but several have been offered, including the Zeigarnik effect of superior memory for unfinished tasks and the Ovsiankina effect of resumption of incomplete tasks. It also could be explained in terms of operant conditioning; specifically, as a kind of "searching behavior" reinforced by the experimenter's comments and by the subject's own successful progress toward a solution. Whatever the explanation, the motivating power evidently does not appear in strength unless the student is required to learn almost completely without help and expends intensive effort over a period of 15 minutes or more.

Meaningful learning. Aside from the advantage the student may come to have academically, he may not benefit from knowing the explanations for rules and procedures he learns, i.e., the pattern of relationships involved. That which is meaningful (understood) may or may not be retained longer and transferred more effectively than that which has been learned by rote. Moreover, superficial efforts to gain understanding after a rule or principle has been memorized may have an inhibitory effect when the student attempts to recall and transfer the original learning. If it is important only that the task be understood (as is most often the case, presumably), the essential relationships may be learned most economically when taught by another person or teaching program, not by process of self-discovery.

REFERENCES

AUSUBEL, D. P. Learning by discovery: Rationale and mystique. *Bull. Nat. Ass. Second. Sch. Principals,* 1961, *45,* 18-58.

BRUNER, J. S. The act of discovery. *Harv. educ. Rev.,* 1961, *31,* 21-32.

CRAIG, R. C. Directed versus independent discovery of established relations. *J. educ. Psychol.,* 1956, *47,* 223-234.

KERSH, B. Y. The adequacy of "meaning" as an explanation for the superiority of learning by independent discovery. *J. educ. Psychol.,* 1958, *49,* 282-292.

KITTELL, J. E. An experimental study of the effect of external direction during learning on transfer and retention of principles. *J. educ. Psychol.,* 1957, *48,* 391-405.

LI, J. C. R. *Introduction to statistical inference.* Ann Arbor, Mich.: Edwards, 1957.

61

Problem-Solving, Past Experience, and Availability of Functions

Per Saugstad and Kjell Raaheim

Saugstad and Raaheim have investigated the role of previous experience in problem-solving behavior. Although the situation is not a "guided learning experience," the reader might wish to examine this paper in relation to Ojemann and Pritchett (24) in terms of the nature of the task. This Norwegian study, drawing on American and German research and published in a British journal, also demonstrates how modern psychological science transcends national borders.

INTRODUCTION

Evidently the higher intellectual processes in man are determined by an enormous amount of past experience. An adequate understanding of

---

From *British Journal of Psychology,* 1960, *51:* 97-104.

thinking, therefore, is inconceivable without a conceptual framework which makes it possible to handle the effect of past experience. American theorists of learning have attempted to provide such a framework in terms of principles of learning derived from a study of some relatively simple situations. However, it seems fair to state that this group of psychologists has neglected the problems of perception. On the other hand, Gestalt psychologists have attempted to make a framework in terms of principles of perception derived from a study of relatively simple perceptual situations. This second group apparently has neglected the problems of past experience and learning (Saugstad, 1958). Evidently thinking must be understood in terms of learning as well as of perception and also in terms of principles for memory. The one-sided approaches to the study of thinking among experimental psychologists may be due to the fact that, during the last generation, the thought processes have relatively seldom been the object of a more direct experimental analysis. Attempts are usually made to account for them in terms of principles derived from studies of simpler situations particularly within learning and perception.

It is a conviction of the writers that not only will an understanding of thinking be increased by a direct experimental attack on problems in this area, but so also will the processes involved in learning and perception. An outline for a framework based upon a more direct attack on problems of problem-solving has been presented elsewhere (Saugstad, 1955, 1958) and will be briefly indicated here.

In the hope of making communication more efficient—an overwhelmingly difficult problem in studies in this area—attention is restricted to problem situations involving concrete objects, such as stones, sticks, ropes, balls, iron nails, boxes, pieces of string, and implements of various sorts (e.g. pairs of pliers, screwdrivers, hammers). In problem situations of such a concrete nature the solution is attained by the use of definite objects in definite ways and, in the attainment of the solution, the objects may be said to serve definite functions. A concrete problem may now be subjected to a logical analysis in terms of these functions. The solution may be conceived of as the arrangement of a number of functions into some definite sequence. In previous experiments (Saugstad, 1955; Saugstad & Raaheim, 1957), the functions available in an individual with regard to a definite object have been assessed by the following simple procedure, called an availability test. The individual is presented with one object at a time and asked, according to some specified rules, to list as many functions for this object as possible. This procedure is carried through without any mention of a definite problem situation.

So far, the term 'function' has been used here in its relatively vague every-day meaning. Instead of attempting to lay down a more formal definition at present, it is thought wiser to attempt to clarify the term slowly as the experimental work progresses. To some extent its meaning is implicitly given by the procedure adopted in the first experiment by Saugstad (1955), but to avoid some misunderstandings (Székely, 1958; Saugstad, 1958) some brief comments on certain aspects will be presented.

Functions may be considered as varying with regard to specificity or generality. One may think of a function as something determined by some very specific context or one may have in mind a large variety of situations. For example, I may use a stone in one specific situation to crack one specified nut, or I may use a stone to crack nuts in general in a variety of situations. This wider generality was what Saugstad (1955) had in mind when he designed his first availability test in such a way that the subjects were asked, first, to state a function in a general form and then to give three concrete examples of the function. The procedure followed in the experiments performed thus rests on the assumption that functions may be perceived or conceived of in a more or less general form. The results of the experiments performed within this framework indicate that, under certain circumstances, functions may be regarded as being relatively independent of the situational context. On the other hand, experiments by Duncker (1945), Adamson (1952), Birch & Rabinowitz (1951) reveal that, under other circumstances, availability of a function may be highly dependent upon the situational context. The degree to which a function is dependent upon the situational context can be clarified only by more systematic empirical research. Attention should be called to the fact that, so far, attempts have been made within this framework to stay away from situations which may be conducive of fixation of a function.

In connection with the problem of specificity or generality of functions, it should be pointed out that the present approach grew out of a dissatisfaction with the currently dominant approaches to learning and problem-solving. On the one hand, there is the approach of Gestalt-psychology where the term 'cognitive or thought structure' so strongly stresses the situational context that the situation can hardly be analyzed into subunits. Evidently human beings and primates generally are, to a considerable extent, able to isolate and manipulate definite aspects of a stimulus situation independent of the context. In this connection, it should be emphasized that, to a large extent, we conceive of concrete objects as discrete. On the other hand, the various approaches in terms of 'stimulus-response' tend to neglect the situational context. This makes it difficult, if not impossible, to predict over a variety of situations what is to be considered as the stimulus. It is hoped that the term 'function' might strike a more happy medium than either the term 'cognitive structure' or the term 'stimulus' as used by a number of behaviorists. It seems considerably easier to isolate and manipulate the referents of the term 'function' than those of the term 'cognitive structure'. Also, in contrast to the term 'stimulus' which tends to be conceived in terms of physical energy and in some cases in terms of physiological processes, the term 'function' will probably cover a larger variety of situations since it has a more definite phenomenological anchoring.

A number of hypotheses with interesting theoretical implications may be formulated with regard to functions. One might, for example, formulate hypotheses relevant to the problem of conditions affecting availability of functions, or hypotheses relevant to the problem of how individuals proceed to arrange a number of functions into the correct sequence for the solution of a problem, or hypotheses concerning the relationship between the ability to produce functions and other types of intellectual abilities. In the experiments performed so far, concentration has been on the following question: will a person who has available in a problem situation the necessary functions for the solution of the problem actually solve this problem? In the exploration of this question the assumption has been made that a function available in the availability test will be available also in the problem situation. Results of the experiments (Saugstad, 1955, 1957; Saugstad & Raaheim, 1957) suggest that the question is to be answered in the affirmative. An individual having available the necessary functions will solve the problem. Some of the theoretical implications of this result have been discussed elsewhere (Saugstad, 1958). Here, it should just be pointed out that it is reasonable to state that an individual, who has available the necessary functions for the solution of a problem, is in possession of the necessary past experience for solving the problem. If experience is determined in this way, the results may be stated as follows: an individual having available the necessary past experiences for the solution of a problem will solve this problem.

The problem of whether or not availability of the necessary past experiences is sufficient for solution is of considerable theoretical interest. Mainly based upon Maier's (1930) well-known pendulum problem, Gestalt psychologists and Gestalt-oriented psychologists (Leeper, 1951) have argued that availability of the necessary past experiences is not sufficient for the solution of a problem. In his experiment, Maier demonstrated to a group of subjects what he considered to be the three essential principles for the solution of the pendulum problem. Since only one in a group of subjects after this demonstration solved the pendulum problem, Maier concluded that availability of the necessary past experiences for the solution of a problem was not sufficient for arriving at the solution. A repetition of this experiment by Saugstad (1957) took steps to make sure that subjects had understood the demonstrations, as purported by Maier and as necessitated by his conclusion. It revealed that close to 100% of the subjects having the principles available solved the problem. The principles involved in Maier's demonstrations were very abstract and this may be the reason why the demonstrations had little or no effect. It might be thought that a demonstration, by way of concrete objects, of the necessary functions for solution might be more effective in conveying to the subjects the necessary past experiences. If the subject having received such a demonstration solved the problem, this would be an effective way of showing that an analysis of problem-solving in terms of functions reveals an important aspect of problem-solving behavior. Such a type of experiment will also have some bearing on the problem of the effects of hints and demonstrations in problem-solving which has been the subject of a number of investigations.

EXPERIMENTAL

Subjects

All the pupils in one class in a secondary school in Oslo served as subjects. The pupils, males aged 17–18 years, were majoring in mathematics and physics. The experiment was planned with two classes, but because the results seemed to be highly reproducible, it was discontinued when one class had been tested.

The Problem

Since the problem has been described in detail by Saugstad & Raaheim (1957), only a rough outline will be given here. The subject is faced with the task of transferring a number of steel balls from a glass into a metal cylinder. The glass containing the steel balls is placed on a movable wooden frame and the metal cylinder to the side of it. The subject has to keep behind a wooden table, or a chalk line, 260 cm. away from the wooden frame. At his disposal for the solution of the problem the subject has four newspapers of a definite size, a length of string, a pair of pliers, five elastic bands, and one iron nail. The correct solution to the problem consists in bending the nail by the pair of pliers into a hook, attaching the nail to the string and throwing the nail into the wooden frame, making it catch onto the frame, and then pulling in the frame on which the glass of steel balls lies. The newspapers are then to be rolled into tubes, which are inserted into each other to form a continuous passage. The newspapers are held together by the elastic bands. Through this tube the steel balls are directed into the cylinder from behind the table.

Procedure

Demonstration of the critical functions. Analysis of the problem had previously shown that all subjects giving the function 'hook' for the nail and the function 'tube' for the newspaper, would solve the problem. It was decided to demonstrate only these two functions, called the critical functions. The demonstrations were carried out by reading the following instructions. 'We will now present you with some objects which all may be used in a number of ways. We want you to give us some examples for the use of the objects.' A newspaper rolled into a tube is handed to the subject. 'This object you may use to conduct something through. Could you give some examples?' A nail bent into a hook is handed to the subject. 'This object may be used to catch things with. Could you give some examples?' When the subject had given as an example in answer to the first question: air, or some other gas, water, sand, balls, pebbles, or some other smaller objects, this was considered a satisfactory indication that he had grasped the meaning of the function stated for the newspaper. When the subject had answered the second question with: angle for fishing, to hang things on, as a hook, or to catch on to wood or some softer material, the example was considered an indication that he had grasped the meaning of the function stated for the bent nail. If the subject did not give any of the examples mentioned he was asked to continue until he had met the criterion as stated.

Design. The subjects were first given the demonstrations as described in one room and then immediately led into another room where they were presented with the problem situation. The control group of an experiment previously reported by Saugstad & Raaheim (1957) served as control group for the present study. This group was presented with the problem without any previous demonstrations. The subjects of this group, forty-five pupils from secondary schools in Oslo, had the same characteristics as the subjects in the experimental group. The subjects of both groups were tested individually.

Subjects were allowed 30 min. on the problem. The demonstrations of the critical functions took from 1 to 3 min.

RESULTS

Table 1 shows the number of subjects, in each group, who had solved the problem after 5, 10, 15, 20, 25, or 30 min. (The percentage number of solutions is shown in brackets, to the nearest whole percent.) The figures show that, at every time interval, the subjects who had received the demonstrations were superior to those who had not. Taking 30 min. as the time limit, solutions were given by nineteen out of the twenty subjects of the experimental group as opposed to only ten out of the forty-five subjects of the control group: this difference is significant at the 1% level (χ^2 corrected for continuity).

Only two subjects in the experimental group spent more than 20 min. on the solution. As noted above the demonstrations took from 1 to 3 min. Therefore, the difference in number of solutions between the experimental and the control group cannot be explained in terms of greater familiarity with the material on the part of the subjects in the experimental group.

The results which seem to be highly reproducible, indicate that subjects from the same

Table 1: *Solutions to the problem of transferring the steel balls*

Group	No. of solutions						Total no. of subjects
	5 min.	10 min.	15 min.	20 min.	25 min.	30 min.	
Exp.	5 (25)	14 (70)	17 (85)	18 (90)	19 (95)	19 (95)	20
Contr.	0 (0)	0 (0)	3 (7)	7 (16)	9 (20)	10 (22)	45

population as the subjects of this experiment will almost always solve the ball problem when the two critical functions have been demonstrated to them.

DISCUSSION

Almost all the subjects who had the two critical functions demonstrated to them solved the problem. This suggests that an analysis in terms of availability of functions uncovers an important aspect of problem-solving behavior. By means of this analysis, it appears to be possible, in a simple way, to provide an individual with the part experiences necessary for the solution of a problem. The analysis of a problem situation in terms of functions makes it possible to predict which aspects in a problem situation should be pointed out to the subjects not solving the problem. The present results thus make it possible to formulate, in a tentative manner, the conditions under which a hint will be effective: a hint will be effective to the extent to which it makes available the critical functions for the solution of a problem.

We will from this angle consider some experiments with concrete objects where hints are introduced. Maier (1931) constructed the problem of tying together two ropes which were hung so far apart that, when an individual was holding one, the other could not be reached. To solve this problem, the subject had to make a pendulum out of one of the ropes by tying an object onto it. In this problem situation, it was found effective to give one of the ropes a push. The push may reasonably be said to make available the function served by the rope. Battersby *et al.* (1953) used the same problem situation, but they gave hints by placing objects which suggested the function of a weight to the subject—thus emphasizing another aspect for the construction of the pendulum. Again it is to be expected, as it transpired, that the hint was effective. Cofer (1951) also used the same situation but had his subjects memorize lists of words before they were presented with the problem. More solutions were produced in a group of subjects who had memorized among other words, rope, swing, and pendulum.

The above formulation for the effect of hints may also help to clarify some of Maier's (1945) reasoning around a more recent experiment. We will concentrate on one aspect of this experiment. Maier presented one group with the problem of constructing a hat rack. This was done by making a longer pole out of two shorter ones, and by placing this longer pole between floor and ceiling and clamping them together by a table clamp. The clamp in this construction served as a hat rack. Another group of subjects were placed before the above-mentioned construction after this had been completed and asked to list possible functions for this construction. It was found that there was a proportionally larger number of subjects in the first group who were able to construct the hat rack than there were subjects in the last group who could suggest the function of a hat rack or some similar function. As we understand him, Maier argues that the situations for the two groups were equivalent and that it is strange, therefore, that the results came out as they did. But the two situations are not equivalent and the results are not at all surprising. Evidently, it is difficult to perceive or conceive of the function of a hat rack in the clamp; few subjects in the second group perceived this function. The first group being assigned the task of constructing a hat rack has, *through the instructions*, received a hint that some object may serve as a hat rack. For this reason, a larger proportion of them finds the function of hat rack for the clamp. It is thus seen that the two groups are treated entirely differently in one crucial aspect. The situation for the subjects receiving the instruction to build a hat rack is similar to that of the subjects in Cofer's (1951) experiment who, before they were presented with the problem of tying the two ropes together, had memorized the list containing the critical words, rope, swing, pendulum. In both experiments, the subjects received important verbal hints concerning the critical function for one crucial object in the attainment of the solution to the problem.

In still another of his ingenious experiments Maier (1933) was able to raise the number of solutions in a group of subjects by giving them a lecture on how to solve problems. The results of this experiment are of course very interesting, even if it is difficult to draw any definite conclusions as to what in the lecture had the effect of helping the subjects to find more solutions. According to the present analysis in terms of functions, it is natural to assume that the lecture helped the subjects find more functions than they otherwise would have found. In view of the content of the lecture this interpretation seems reasonable.

The present results also serve as a check on one of the assumptions underlying the conclusion drawn from previous experiments by Saugstad (1955) and Saugstad & Raaheim (1957) that a subject having available the necessary functions in the problem situation will solve the problem. Some of the main assumptions underlying this conclusion have recently been discussed elsewhere by Saugstad & Raaheim (1959). One of these had a bearing on the fact that the availability tests might have a selective effect. The subjects who, according to the availability tests, were credited with the necessary functions might be superior in general intelligence or in some problem-solving ability. Since they were superior in one or both of these respects they might be superior in combining the necessary functions into the correct sequence involving the solution. Thus, there might possibly exist a group of subjects having the necessary functions available in the problem situation, but not being able to combine the functions into the correct sequence. Previous analysis (Saugstad & Raaheim, 1957) makes it probable that there exists *no* such group of subjects. The results of the present experiment give a further check on this assumption. In the experimental group, close to 100% solved the problem as against 20% in the control group. The problem solvers in the first group can, therefore, not reasonably be considered superior to the problem solvers of the second group with regard to a possible problem-solving ability or to a possible general intelligence. When almost all subjects of the experimental group solved the problem the reason is most likely that the main determiner of success in problem-solving is availability of functions. However, it should be noted that the combination of the functions for the subjects in the experimental group might be somewhat easier, since the previous demonstrations might have emphasized the two critical functions and thus made easier the combining of them into the solution. Yet it seems safe to conclude that there were still a number of possible functions which might be combined. If the ability to combine functions or some sort of parts was an important factor in determining success in problem-solving one would expect it to manifest itself also in the problem situation of this experiment.

REFERENCES

ADAMSON, R. E. Functional fixedness as related to problem-solving: a repetition of three experiments. *J. exp. Psychol.*, 1952, *43*, 288–291.

BATTERSBY, W. S., TEUBER, H. L., & BENDER, M. B. Problem-solving behavior in men with frontal or occipital brain injuries. *J. Psychol.*, 1953, *35*, 329–351.

BIRCH, H. G., & RABINOWITZ, H. S. The negative effect of previous experience on productive thinking. *J. exp. Psychol.*, 1951, *41*, 121–125.

COFER, C. N. Verbal behavior in relation to reasoning and values. In H. Guetzkow (Ed.), *Groups, leadership, and men*. Pittsburgh: Carnegie Press, 1951.

DUNCKER, K. On problem-solving. *Psychol. Monogr.*, 1945, *53*, No. 5 (Whole No. 270).

LEEPER, R. Cognitive processes. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.

LEWIN, K. Behavior and development as a function of the total situation. In L. Carmichael (Ed.), *Manual of child psychology*. New York: Wiley, 1946.

MAIER, N. R. F. Reasoning in humans: I. On direction. *J. comp. Psychol.*, 1930, *10*, 115–143.

MAIER, N. R. F. Reasoning in humans: II. The solution of a problem and its appearance in consciousness. *J. comp. Psychol.*, 1931, *12*, 184–194.

MAIER, N. R. F. An aspect of human reasoning. *Brit. J. Psychol.*, 1933, *24*, 144–155.

MAIER, N. R. F. Reasoning in humans: III. The mechanisms of equivalent stimuli and of reasoning. *J. exp. Psychol.*, 1945, *35*, 349–360.

SAUGSTAD, P. Problem-solving as dependent on availability of functions. *Brit. J. Psychol.*, 1955, *46*, 191–198.

SAUGSTAD, P. An analysis of Maier's pendulum problem. *J. exp. Psychol.*, 1957, *54*, 168–179.

SAUGSTAD, P. Problem-solving and availability of functions: A discussion of some theoretical aspects. *Acta Psychol.*, 1958, *13*, 384–400. *Nord. Psykol.*, 1958, *10*, 216–232.

SAUGSTAD, P., & RAAHEIM, K. Problem-solving and availability of functions. *Acta Psychol.*, 1957, *13*, 263–278. *Nord. Psykol.*, 1957, *9*, 205–220.

SAUGSTAD, P., & RAAHEIM, K. Problem-solving and availability of functions in children. *Acta Psychol.*, 1959, *16*, 45–58. *Nord. Psykol.*, 1959, *11*, 45–58.

SZÉKELY, L. Some comments on problem-solving availability and test-magic. *Acta Psychol.*, 1958, *14*, 152–157. *Nord. Psykol.*, 1958, *10*, 108–113.

S.C.E.R.T., West Bengal
Library
Calcutta
25/3, B. C. Ro